Università
Ca'Foscari
Venezia

Corso di Laurea magistrale
in Economia e Finanza

Tesi di Laurea

Ca' Foscari
Dorsoduro 3246
30123 Venezia

Fractal and multifractal models for price changes

An attempt to manage the Black Swan

Relatore
Ch. Prof. Marco Tolotti

Laureando
Federico Maglione
Matricola 821095

Anno Accademico
2012 / 2013

A Zia Bella

Fractal and Multifractal Models for Price Changes

An attempt to manage the Black Swan

Federico Maglione

Contents

Introduction

> The point of departure is that financial prices, including those of securities, commodities, foreign exchange or interest rates, are largely unpredictable. The best one can do is to evaluate the odds for or against some desired or featured outcomes, the most extreme being "ruin". Those odds are essential to the scientist who seeks to understand the *mechanism of financial markets* and other aspects of the economy. They will also be used as *inputs* for decision concerning economic policy or changes in institutional arrangements. To handle all those issues, the first step - but far from the last! - is to represent different instances of price variation by suitable random processes.
>
> Mandelbrot, B.

At first glance, the expressions "suitable" and "processes" may surprise. In large part of academic papers and practice, price changes are described by a well-know stochastic process, that is a Browninan Motion (or its discrete version which is named Random Walk[1]). It is very similar and connected with the coin tossing phenomenon: Prices can move either up a bit or down a bit following the toss of a coin, being their movements absolutely independent. Most of the financial economists (intended in the largest sense) do believe that there is no alternative to this model of price variation.

To the contrary, we will follow the effort made by renowned mathematicians, physicists, and even economists – the notorious econophysicists –, whose forefather can be doubtless considered the French mathematician Benoît Mandelbrot, who tried to change that mindset. In fact, the aim of this work is to show that the word "random" has a far wider meaning, allowing the coin-tossing model to be replaced by alternative ones, more suitable to descrive the reality of financial markets.

And with a Mandelbrot's vivid metaphor this concept can be fully understood: When a ship was build to navigate placid lakes by fair weather, sending it across the ocean in typhoon season may be a very dangerous action. Similarly, using the Brownian motion to describe financial prices may

[1] The matter is little more sophisticated. As a matter of fact the Brownian Motion can be constructively "built" starting from a Random Walk: Taking the limit process of the latter and normalizing the time interval, we get a Brownian motion.

be beloved by mathematicians and economists, but at the same time it denies the existence of hurricanes. The coin-tossing model exemplifies a state of randomness that Mandelbrot addressed as *mild*. However financial markets, like rough water, are driven by a *wild* form of randomness[2].

Hence, policy makers and practitioners in finance should be far more demanding. For them, the inappropriateness of Brownian motion may become a real concern. This model was proposed to describe the equilibrium of the particles of a fluid. Nevertheless, in Finance the characterization of market structures does not lie in the equilibrium, but rather in the violent turbulence. According to Brownian motions, "financial hurricanes" never ought to happen.

Moreover, ship-builders and ship owner can neither forecast when a hurricane is going to happen nor how much aggressive it shall be, but exact prediction is not needed: The knowledge that hurricane will happen, together with the realistic valuation of the corresponding probabilities, permeates ship-building, ship ownership and navigation.

The Brownian motion model served as foundation for the theory of Classical Finance. A blatant reason can be traced in being the least complex stochastic process to guide a random phenomenon (as price changes are). It possesses several attractive properties and it is easy to manipulate in order to get advanced results, especially in the matter of the option pricing. However it also owns few limits, maybe too many and really significant. First of all it predicts *normal* variation, that is extreme events, like a daily decrease about 20%, are very unlikely. In other cases the event has a so little odd to be considered as impossible. But such drops happen and with a frequency extremely higher than in the normal case. To take the most famous (and tragic) tumble of the history, the Dow Jones Industrial Average dropped by 29.2% in one day. The odd of such a variation, according to Brownian motion and hence normal distribution, is less than 10^{-50}, being the likelihood so little to necessarily be considered as an impossible event.

Everybody know that financial market are risky. However a in-depth study of this concept – the risk – can lead to a new comprehension of our (financial) world, hoping to be able to face and control it quantitatively. For more than a century, financial experts and economists have struggled to analyze the risk involved into financial markets, to explain and quantify it, and to gain profit from it. In our opinion, most part of theorists has explored a wrong avenue and, moreover, even when they recognized it, they proceeded as nothing had happened. During the last years, they have created a mass of complex mathematic tools in order to evaluate risk which has been adopted by the most famous financial institutions. They have tried to combine portfolios of assets, attempting to reach a fixed trade-off between

[2]For a formal definition of *states of randomness* see [27].

risk and return. On the other hand, the dramatic falls occurred in the eighties, in the nineties and, again, in the two thousands have been coercing a change of mind both to financial experts and economists. If profitability and risk have to work proportionally, the usual computations are wrong since the denominator, the risk, is *larger* that how it is commonly evaluated. A more careful risk's measurement, together with a new comprehension of the way in which risk points the markets, are the (maybe pretentious) aims of this thesis.

A new vision of risk will be given by the means of different (and perhaps sophisticated) stochastic processes, with the great contribution of Fractal Geometry. This branch of mathematics will enable us to give new order to those phenomenon which seem to be just chaotic, without a underlying leading structure.

At the same time, we have to remark that several generalizations of Bachelier's random walk model have been developed, based on long-memory stochastic processes and discontinuous trajectories. The first class was explored by Mandelbrot himself, leading to Fractional Browninan Motions. On the other hand, the second eventuality was taken into account by the French Paul Lévy, introducing the general class of Lévy Processes and then Pure Jump Processes, which constitute a subclass of the former.

In the first chapter we will introduce the Browninan Motion and two alternatives which, in different fashion, allow the existence of typhoons and hurricanes. In the second chapter, we will focus on those tools needed to deeply understand the misbehaviour of financial markets that are already available, but little-known: They are those of fractal and multifractal geometry. Only in the third chapter we will analyze the *multifractal* process, which differs in "qualitative" ways if compared with the previous ones. It will also allow us to summarize and merge all those features required for an authentic representation of financial markets' complexity.

Critical aspects of Classical Finance

Risk can be handled in different ways. In the financial markets' context, the oldest technique, that is also the most elementary one, is represented by the *fundamental analysis*. If the price of stock increases, the causes are traced inside the corporation or with facts concerning the industry in which the latter operates. With a deepened study, we should be able to foresee the next move of the stock. The focus is on the causal nexus: The price of a stock, bond, derivative, or currency varies since a particular event, coming from the outside of the market, has happened. The wheat price grows since a heat wave desiccates the producer country. Financial press takes benefit from it, since the news are sold, and financial corporations

make a business in categorizing the information: Thousands of analysts, operating at a macroeconomic or at a microeconomic level, following a "top-down" or a "bottom-up" approach, work everyday looking for that piece of information able to make money. On the other hand, financial authorities use those analyses for both internal and external disclosure, and impose laws on what information companies are required to publish (the balance sheet, the profit and loss account, budget plans, the deliberations of the board of directors). The premise of fundamental analysis is that if we know the cause, we can forecast the event and then control the risk involved.

Of course the matter is absolutely more complex. In the real world, causes are frequently not clear at all. Critical information is often unknown or unknowable. It can be hidden or misunderstood. Only *a posteriori* facts become evident and a fundamental approach can be tracked efficiently. Thus, how a financial strategy can be exclusively based on the principle of knowing earlier and more than other else?

The answer were given by the financial industry with another tool. The style of analysis immediately subsequent to the fundamental one is represented by the *technical analysis*. It consist of identifying structures, configurations and schemes (real or presumed) in charts of prices, trading volumes or other indexes/functions based on them, looking for a signal or evidence that suggest us to sell or buy the asset in point. The language of the chartists (the analysts that base their decisions and forecasts on the graphical aspect of the patterns of prices) is extremely motley: "Head and shoulder", flags and triangles are those signals they seek for. But such an approach is mostly used in the currency markets. There, the main firms operating on the exchange rate markets employ a sizable quantity of technical analysts to find "points of control", "trading range" and many other data configurations. And inside the logic of markets, with their mechanisms like a mirror-maze, *sometimes* chartists may be right. The exchange rate between two currencies can really come up to the target foreseen by the technical analysts, and then come back like a ball bouncing of the walls or speed up their grow as an invisible barrier were torn down. But this is like an illusion, or maybe a swindle: Everybody know that all the others agents in the market know the existence of the intervention and they act accordingly. It is crazy that huge quantities of money may change owner according to such financial "astrology". As previously stated, it can work sometimes, but it cannot be ascribed as a bedrock on which to build a global system of risk controlling.

At a later stage, what is commonly named into Faculties of Economics by *modern finance* was developed, emerging from rigorous mathematics and statistics. The basic notion is that prices are unpredictable, but their movements can be described by the mathematical laws of randomness. Thus, risk becomes measurable and controllable.

Such researches started with Louis Bachelier at the beginning of the last century, introducing a new model commonly addressed by "random walk".

Its premise was that prices could go up and down with the same odd, like a flip of coin shows one face or the other. And hence the variation is measurable: For the most part (the 68 per cent), price variations are little deviations from the mean less than one standard deviation; in the 95 per cent of cases, the variations should be less than two standard deviation; in the 98 per cent of cases less than three standard deviations. Eventually, large deviations ought to be in a very limited number. If we make an histogram of all the price variations, we should see a bell curve: Many little variations are gathered in the centre of the bell, the rare large ones in the tails. For mathematicians the bell curve is *terra cognita*, insomuch as they called it *normal* (suggesting that all the other form are abnormal) or Gaussian, taking the name of its discoverer.

A more general modification of Bachelier's model is called *efficient-market hypothesis*, which were introduced by a Mandelbrot's doctoral student, Professor Eugene Fama (see [17] and [31]). It states that, in a perfect market, prices reflect all publicly available information and that prices instantly change to reflect new public information. A possibility to illustrates this concept is that yesterday's variations do not influence today's ones; subsequently the latter does not influence tomorrow's variations. Every price variation is *independent* by the preceding one.

Based on those theories, economists developed a mass of sophisticated tools in order to analyze financial markets, measure the variance and the beta-coefficient of an asset, thus categorizing portfolios according to they riskiness. According to the mainstream, a fund manager can create an *efficient* portfolio with a target rate of return, given a desired/affordable level of risk. Moreover, employing financial instruments such as derivatives, we can also gain more without taking upon ourself additional units of risk. It is an elegant theory, but, unfortunately, it is inaccurate. Booms and crashes of the last years witness the inadequacy of financial orthodoxy, which found its principles upon two ideas of Bachelier: The statistical independence of price variations and their normal distribution.

First of all, price variation are non independent. Many researches of the last decades indicate that usually financial time series possesses a sort of *memory*, different from those form of periodicity involved into standard economic cycles. What happens today will influence almost surely what is going to happen tomorrow. The memory we will deal with in this thesis originates from a different statistical relationship, of fractal style, that is a *long-range memory*. Moreover, different price time-series exhibit different intensities of memory. The reason underlying this phenomenon is not sure, but we may suppose that the market is not efficient in the informational sense, and uses more time in absorbing and evaluating information. When (bad) news arrive, there are ready investors that react immediately, but there may be others that, because of their different financial aims or investment horizons, respond after a week, a mouth or a year. Whatever the reason may

be, this phenomenon exists and contradicts the random walk assumptions.

Secondly, it is universally recognized that price variations do not follow the Gaussian curve. The extreme daily variations of the Dow Jones Industrial Average, the ones that are in the *tails* of the distribution, are too many if compared to the ones forecast by the normal density.

The following figure show how much these statements are true. According to the classical finance, Dow Jones's extreme variation (the ones greater than the 7%) should occur once every 300.000 years, but during the twentieth-century 48 of such days were registered; the same is found for the FTSE MIB, that is the most representative index of the Italian stock market. We compared the index's return with the ones generated by a Gaussian white noise (see Section 1.2), which is the theoretic consequence of a Brownian motion leading the price dynamics. All the graphs (time-series plot, histogram and normal Q-Q plot) show that the financial returns exhibit more extreme values, having so fat tails. Maybe the last pattern show better how much the sample deviates from normality: In that case, the empirical quantiles are expected to place themselves on a straight line; otherwise, if at the edges they are not linear, it means that the phenomenon is much more erratic.

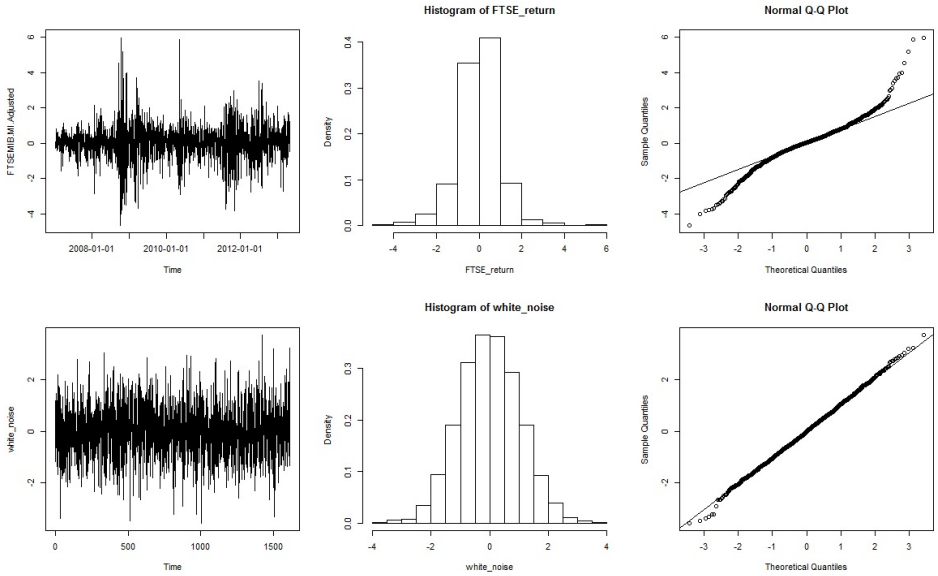


Figure 1: FTSE MIB log-returns from 01/01/2007 to 05/01/2013

An explanation of such a behaviour can be ascribed to the fact that price variation could follow a power-law distribution rather than the bell curve. We shall see that there is a class of random variables whose tails decays in a way that allows the presence of large variations. Subsequently, in place of variance and beta-coefficients accepted by classical finance we will introduce different models based on new variables: The H exponent for

the price dependence and the α parameter which characterizes volatility and heaviness of the distribution's tails.

Today those objections do not fall on deaf ears anymore; many economists and financial analysts have described such misbehaviour in several publications. On the other hand, old models are surprisingly die-hard. The formulas of classical finance, the ones of Bachelier and his heirs about how to choose an investment portfolio, how to determine the present value of a company's real asset, how to judge the riskiness of a security, have taught in the Faculties of Economics worldwide and constituted the linchpin of the qualifying examinations to become a chartered financial analyst. Although to a less extent, they also are part of the tools used by practitioners. As an instance, the notorious Black-Scholes formula (which will be discussed in Appendix D) was for long the *gold standard* to evaluate the stock option of many corporation managers. Only in 2004, after the Internet speculative bubble, the American financial authority officially allowed different evaluation models.

However, especially financial industry is one of the most demanding and pragmatic, having never intended to loose money because of the inaccuracy of a formula. So it invested on research and many new efficient models are available. And the same pragmatism belongs to Central Banks directors. After having accepted for years those old models, in 1998 they began to request the implementation of new ones, so that banks were able to estimate more precisely and realistically their risks. Eventually, as everyone with good sense should admit, this opportunity failed, producing a Revised International Capital Framework (commonly named Basel II) which allowed bank to become even more hazardous in assuming risks. Not to mention the fact that the real progresses in creating more efficient models was very meager, being the latter based always on normal distribution, copulas and so on. (QUI).

Changing science approaches, ideas and models is not so obvious as it should be expected, but maybe in the economic field it becomes even more difficult. Frequently economists turn out to be very jealous of their "creatures", totally blind to those changes that could really make the difference, especially if they come from a non-economic or financial contest (like, despite the name, econophysics). With available advances in modeling, we need to understand that some older models may not be relevant in a particular situation, and yet some modeling continues to use older, less relevant approaches.

To tell the truth, there has been niche efforts in new direction (different to the ones discussed in this work) which, starting afresh, developed a *behavioural* theory of financial decision making. Here investors are supposed not to be perfectly rational being, but are considered as agents that use information in different ways producing different effects. Moreover, *genetic algorithms* and *neural networks* were implemented with the hope that a silicon intelligences were able to recognize profitable markets' configuration

where carbon-based ones did not succeed.

Today, looking into the future Basel III, another chance – perhaps the last one – is given in order to prevent financial crises. A new *post-modern finance* can raise, eventually producing a virtual success.

Chapter 1

Models to describe price changes

This first chapter aims at displaying the best-known model for price variations, namely **Standard Brownian Motion**. It was firstly introduced in finance by the French mathematician Louis Bachelier in 1900 as background of his doctoral thesis "Théorie de la Spéculation", supervised by professor Henri Poincaré. He was the first one to suggest that Probability Theory could be applied to the study of financial markets. As a matter of fact, he is considered a pioneer in the study of financial mathematics.

Five years before Albert Einstein and Norbert Wiener, Bachelier described all the essential features of Standard Brownian Motion introducing the notion of stochastic process. The French mathematician dealt with financial price variations harking back to a physical model proposed by the Scottish botanist Robert Brown from whom the name of the motion comes from.

In 1829, the latter while examining at the microscope few dust particles immersed in water saw them blazing a random path. This fundamental hint allowed Bachelier to model stock price fluctuations. Only in 1905 Einstein, independently, gave a theoretical explanation to the phenomenon reported by Brown, suggesting that the movement of particles could be described hypothesizing their jumps due to stochastic collisions between molecules.

Only later, in the fifties, economists began to take into account the Random Walk model – which can be conceived as the discrete version of the Brownian Motion – and its implications concerning Bechelier's work about price changes. Since then Bechelier's ideas and his model became popular even with economists. Another scientist which contributed formalizing the Random Walk Hypotheses was the physicist Maury Osbourne, applying them to the financial markets behaviour as well. Particularly he elaborated a random process according to which price variations corresponded to the motion of a particle inside a fluid, namely a Brownian Motion. Since in that

way price changes were considered being independent, in addiction they were expected to be Gaussian, with stable mean and finite variance, due to Cental Limit Theorem.

Nevertheless, just with the introduction of an axiomatic Probability Theory by Andrej Kolmogorov in 1930, Standard Brownian Motion could be completely understood as a time-continuing stochastic process (named by him Wiener spiral, today Wiener process).

The following step of this chapter is to show two other models which relaxes several hypotheses of Standard Brownian Motion. The first will be the so-called **Fractional Brownian Motion**, which is usually thought to be introduced by Benoît Mandelbrot, but its origins can be ascribed to Kolmogorov himself, who used it in 1940 for a different aim and contest. Developing the turbulence theory of Reynolds number, he introduced the notion of field with statistically isotropic and homogenous increments. The usage of this different process is justified when price variations are assumed to be *dependent* on the previous ones, having so infinite memory; indeed Standard Brownian Motion is not able to capture this kind of behaviour. Of course the latter is a particular case of the more general Fractional Brownian Motion.

For Fractional Brownian Motion development, the work of the British hydrologist Harold Edwin Hurst remains crucial. From 1907 to the early fifties, he worked on the Nile dam building project, dealing with water supply management problem. He asserted that water reserves did not have to overflow their banks nor had to deplete in order to meet the water demand. Attempting to define a mathematical model able to simulate that virtual problem, Hurst hypothesized the stream of water towards the river basin as an uncontrollable part of the whole system following a random walk *ante litteram*. If it had been confirmed, the fluctuation range should have grown proportionally to the square root of the time between two measurements. But Hurst's empirical researches on the overflows of the Nile contradicted the absence of memory involved in random walk hypotheses. So Hurst introduced, heuristically, the concept of persistent random walk: After a increase of the water reserve level, it was more likely to observe an additional increase rather than a drop. To catch that behaviour Hurst introduced a measure of dependence, H, called Hurst exponent. The definition of Fractional Brownian Motion is based on this quantity.

Eventually, the last topic of the chapter will be the presentation of a class of different random variables which result to be the *only one* possible limiting distributions for a sum of independent, identically distributed, random variables. Furthermore, these distributions posses other attractive properties with respect to financial prices: They have a asymptotically scaling behaviour in the extreme tail areas and are invariant under addiction. Those are the **L-stable distributions**. The derivation of most of the im-

portant properties of them is due to the mathematician Paul Lévy (so the name L-stable) in 1925, but a rigorous and compact mathematical treatment of the statistical theory can be found firstly in Gnedenko and Kolmogorov in 1954 and in Mandelbrot later. A more descriptive exposure of the statistical theory was given by the economist Eugene Fama in 1965 (see [15]). Even for L-stable distributions, Standard Brownian Motion can be considered as a particular case.

1.1 Standard Brownian Motion (SBM)

We now define the *Standard Brownian Motion*[1] (briefly indicated as SBM), which constitutes a preeminent element of stochastic processes theory and has been widely used in quantitative finance to describe price variations. As a matter of fact, it can be considered as the most important continuous-time stochastic process, like the normal distribution is seen as the "passe-partout" one in many context, when referring to random variables. Brownian motion can be also seen as the time-continuous version of a discrete random walk with Gaussian increments: It is indeed possible to prove that SBM can be obtained as an appropriate limit of whatever random walk with variance-finite increments. In addiction, it is even called Wiener Process.

We can now define and study SBM.

Definition 1.1.1 *Given a probability space* $(\Omega, \mathcal{F}, \mathbf{P})$*, every real stochastic process* $\{B(t)\}_{t\in[0,+\infty)}$ *defined on it is said* standard Brownian motion*, if it satisfies the following properties:*

(a) $B(0) = 0$ *almost surely;*

(b) *for every* $0 = t_0 < t_1 < t_2 < ... < t_n < +\infty$ *all the increments* $B(t_1) - B(t_0)\,,\, B(t_2) - B(t_1)\,,\, ...\,,\, B(t_{n-1}) - B(t_n)$ *are independent;*

(c) *for every* $0 \leq s < t$*, all increments* $B(t) - B(s)$ *are zero-centred stationary Gaussian with null mean and variance equal to the width of the time interval, that is:*

$$B(t) - B(s) \overset{d}{\sim} \mathcal{N}(0, t-s)\,; \tag{1.1}$$

(d) *there exists* $A' \in \mathcal{F}$*, with* $\mathbf{P}(A') = 1$*, such that for all* $\omega \in A'$ *the map* $t \to B(t,\omega)$ *is continuous.*

[1] This process is addressed by many authors simply as *Brownian Motion*. The clarification "standard" is because we will set $\mathbb{E}[B(1)^2] = 1$. For more details, see Definition 1.1.13.

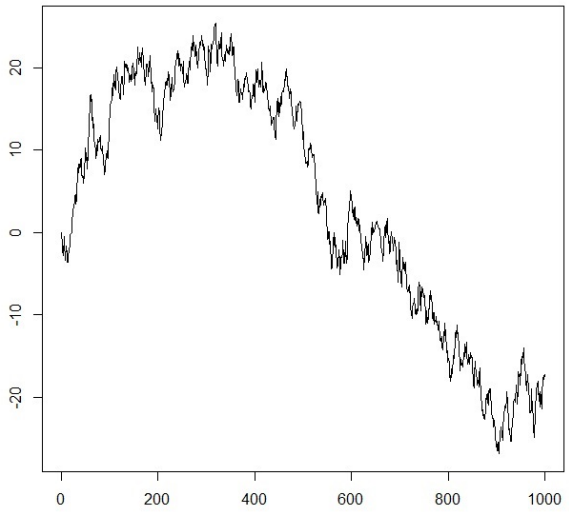


Figure 1.1: Discrete version of a standard Brownian motion (random walk)

Before analyzing some features related to this class of processes, it may be informative to mention that the condition **(c)** could be replaced by a more general one. In fact, it can be relaxed requiring only an identical distribution of stationary increments; this one leads to a process $\{a \cdot B(t) + b \cdot t\}_{t \in [0,+\infty)}$, where $\{B(t)\}_{t \in [0,+\infty)}$ is the standard Brownian motion as defined below and $a, b \in \mathbb{R}$. In other words, for less than a scaling factor and an addiction of a linear function, the standard Brownian motion is the *only one* null in the origin, with independent and stationary increments and continuous paths stochastic process.

The following part of this section is devoted to the mathematical statement of some really important results about SBM. We will get proof of some of them. The first is the existence Wiener theorem, whose first proof was given by Wiener himself in 1923 [41]. Other ways of proving are available and the most famous are the one of Lévy and the more general one due to Kolmogorov. However, we will only report it.

Theorem 1.1.2 (**Wiener theorem**) *There exists a standard Brownian motion on some probability space.*

That being stated, the following characterization of SBM motion is crucial.

Theorem 1.1.3 *A real stochastic process $\{B(t)\}_{t \in [0,+\infty)}$ is a SBM if and if only is a Gaussian process with mean $\mu_B(t) = 0$, autocovariance function $\mathcal{C}_B(t,s) = \min\{t,s\}$ and continuous paths almost surely.*

Proof (1): We will prove that, if $\{B(t)\}_{t \in [0,+\infty)}$ is a SBM, therefore $\mu_B(t) = 0$ and $\mathcal{C}_B(t,s) = \min\{t,s\}$.

For all $0 = t_0 < t_1 < ...t_n < +\infty$, we set

$$Y_0 = B(t_0)\,,\, Y_1 = B(t_1) - B(t_0)\,,\, ...\,,\, Y_k = B(t_k) - B(t_{k-1})$$

with $k = 1, 2, ..., n$. Due to Definition 1.1.1, we have a.s. $Y_0 = 0$, $Y_1 = B(t_1)$, ...,$Y_k = B(t_k) - B(t_{k-1})$, thus the vector $(Y_0, Y_1, ..., Y_k) \in \mathbb{R}^{k+1}$ is a random Gaussian vector. It follows that even the vector $(B(t_0), B(t_1), ..., B(t_k))$ is Gaussian since it is a linear transformation of the previous one. This shows that $\{B(t)\}$ is a Gaussian process. In addiction, due to property **(c)** of the same definition, we have that the random variable $B(t) - B(t_0) = B(t)$ is normally distributed with zero mean and variance equal to the time-expired, that is:

$$B(t) \overset{d}{\sim} \mathcal{N}(0, t)\,. \tag{1.2}$$

Thus $\mu_B(t) = \mathbb{E}[B(t)] = 0$ for all $t \geq 0$. Regarding the autocovariance function (equal to the autocorrelation one in that case), letting $s \leq t$ we have:

$$\begin{aligned}
\mathcal{C}_B(t,s) &= \mathbb{E}[B(t) \cdot B(s)] = \mathbb{E}\big\{[B(t) - B(s) + B(s)] \cdot B(s)\big\} = \\
&= \mathbb{E}\big\{[B(t) - B(s)] \cdot B(s)\big\} + \mathbb{E}\left[B(s)^2\right] = \\
&= \mathbb{E}\big\{[B(t) - B(s)] \cdot [B(s) - B(0)]\big\} + \mathbb{V}[B(s)] = \\
&= s
\end{aligned}$$

because of the independence of the increments. On the other hand, if $s > t$, we have:

$$\begin{aligned}
\mathcal{C}_B(t,s) &= \mathbb{E}[B(t) \cdot B(s)] = \mathbb{E}\big\{B(t) \cdot [B(s) - B(t) + B(t)]\big\} = \\
&= \mathbb{E}\big\{[B(s) - B(t)] \cdot B(t)\big\} + \mathbb{E}\left[B(t)^2\right] = \\
&= \mathbb{E}\big\{[B(s) - B(t)] \cdot [B(t) - B(0)]\big\} + \mathbb{V}[B(t)] = \\
&= t\,.
\end{aligned}$$

Thus, in a more compact way we can write:

$$\mathcal{C}_B(t,s) = \begin{cases} s & \text{if} \quad s \leq t \\ t & \text{if} \quad s > t \end{cases} = \min\{s, t\}\,. \tag{1.3}$$

Proof (2): Finally we prove that, if $\{B(t)\}_{t \in [0,+\infty)}$ is a Gaussian process with $\mu_B(t) = 0$ and $\mathcal{C}_B(t,s) = \min\{t, s\}$, so it is a SBM. The property **(a)** of the definition is trivial since $B(0) \overset{d}{\sim} \mathcal{N}(0, 0)$, so the distribution collapsing in the origin, that is $B(0) = 0$ a.s.. To prove the property **(c)**, we see that since the process is Gaussian, the variable $B(t) - B(s)$ is Gaussian as well, with mean:

$$\mathbb{E}[B(t) - B(s)] = \mathbb{E}[B(t)] - \mathbb{E}[B(s)] = \mu_B(t) - \mu_B(s) = 0$$

and variance (if $s \leq t$):

$$\begin{aligned}\mathbb{V}[B(t) - B(s)] &= \mathbb{V}[B(t)] + \mathbb{V}[B(s)] - 2 \cdot \mathbb{E}[B(t) \cdot B(s)] = \\ &= \mathcal{C}_B(t,t) + \mathcal{C}_B(s,s) - 2 \cdot \mathcal{C}_B(t,s) = \\ &= t + s - 2 \cdot s = t - s\,.\end{aligned}$$

Eventually, to prove the independence of the increments, since they are Gaussian, it is sufficient to prove that they are two-by-two uncorrelated. Letting $1 \leq i < j \leq k$, we have[2]:

$$\begin{aligned}&\mathbb{C}\mathrm{ov}[B(t_j) - B(t_{j-1}), B(t_i) - B(t_{i-1})] = \\ &\quad = \mathbb{C}\mathrm{ov}[B(t_j), B(t_i)] + \mathbb{C}\mathrm{ov}[B(t_{j-1}), B(t_{i-1})] - \mathbb{C}\mathrm{ov}[B(t_j), B(t_{i-1})] - \\ &\qquad - \mathbb{C}\mathrm{ov}[B(t_{j-1}), B(t_i)] = \\ &\quad = \mathcal{C}_B(t_j, t_i) + \mathcal{C}_B(t_{j-1}, t_{i-1}) - \mathcal{C}_B(t_j, t_{i-1}) - \mathcal{C}_B(t_{j-1}, t_i) = \\ &\quad = t_i + t_{i-1} - t_{i-1} - t_i = 0\end{aligned}$$

since $t_{i-1} < t_i \leq t_{j-1} < t_j$.

□

The latter theorem shows SBM is a stochastic process with *independent and stationary increments* (and so it cannot be stationary itself) with $\mu_B(1) = 0$ and $\sigma_B^2(1) = 1$. It is even possible to prove that all the process moments higher than the second order are such that $(k > 2)$[3]:

$$\mathbb{E}\left\{[B(t) - B(s)]^k\right\} = o(t - s)$$

and hence

$$\mathbb{E}\Big[B(t)^k\Big] = o(t)\,.$$

Another important result is about some *invariance* properties of this stochastic process; since the proof is quite trivial, we will not exhibit.

Proposition 1.1.4 *If $\{B(t)\}_{t \in [0,+\infty)}$ is a standard Brownian motion, the following processes $\{X(t)\}_{t \in T^*}$ are SBM as well:*

(a) $X(t) = -B(t)$, *with* $T^* = [0, +\infty)$ (space reflection)*;*

(b) $X(t) = B(t_0 + t) - B(t_0)$, *with* $T^* = [0, +\infty)$, $\forall t_0 \geq 0$ (time translation)*;*

[2] Remember the property on the covariance of linear combinations of random variables $\mathbb{C}\mathrm{ov}(X \pm Y, Z \pm W) = \mathbb{C}\mathrm{ov}(X, Z) + \mathbb{C}\mathrm{ov}(Y, W) \pm \mathbb{C}\mathrm{ov}(X, W) \pm \mathbb{C}\mathrm{ov}(Y, Z)$.

[3] Given $g(x) \neq 0$, we say $f(x) = o\,(g(x))$ in a neighborhood x_0, if $g(x)$ grows much faster than $f(x)$ as $x \to x_0 \in \mathbb{R}^*$. So we have:

$$\lim_{x \to x_0} \frac{f(x)}{g(x)} = 0\,.$$

In the case above we have $t - s \to 0$.

(c) $X(t) = B(t_0 - t) - B(t_0)$, *with* $T^* = [0, t_0]$, $\forall t_0 \geq 0$ (time reflection);

(d) $X(t) = \frac{1}{\sqrt{c}} \cdot B(c \cdot t)$, *with* $T^* = [0, +\infty)$, $\forall c > 0$ (diffusive rescaling);

(e) $X(t) = t \cdot B\left(\frac{1}{t}\right)$ *if* $t > 0$ *and* $X(0) = 0$ *if* $t = 0$, *with* $T^* = [0, +\infty)$ (time inversion).

The following theorem is the counterpart of the Law of Large Number on random variables.

Theorem 1.1.5 (**Law of large numbers for SBM**) *If* $\{B(t)\}_{t \in [0,+\infty)}$ *is a standard Brownian motion, it follows:*

$$\lim_{t \to +\infty} \frac{B(t)}{t} = 0 \quad \text{a.s.} \tag{1.4}$$

Proof: We define

$$X(s) = \begin{cases} s \cdot B\left(\frac{1}{s}\right) & \text{if} \quad s \geq 0 \\ 0 & \text{if} \quad s = 0\,. \end{cases}$$

Due to Proposition 1.1.4, letter **(e)**, that process is a standard Brownian motion itself. From the definition of SBM, it follows a.s. $\lim_{s \to 0} X(s) = 0$. Setting $t = \frac{1}{s}$, we can write:

$$\lim_{t \to +\infty} X\left(\frac{1}{t}\right) = 0 \qquad \Longrightarrow \qquad \lim_{t \to +\infty} \frac{B(t)}{t} = 0\,.$$

□

Next, we present two joined theorems which describe the global and local behavior of the paths of the standard Brownian motion. We will see that they are extremely erratic, being even nowhere differentiable. For their proof see [32].

Theorem 1.1.6 (**Global law of the iterate logarithm**) *If* $\{B(t)\}_{t \in [0,+\infty)}$ *is a standard Brownian motion, the following limits hold almost surely:*

$$\limsup_{t \to +\infty} \frac{B(t)}{\sqrt{2 \cdot t \cdot \ln|\ln(t)|}} = 1 \quad \textit{and} \quad \liminf_{t \to +\infty} \frac{B(t)}{\sqrt{2 \cdot t \cdot \ln|\ln(t)|}} = -1\,. \tag{1.5}$$

From this theorem, it follows that, for all given $\varepsilon > 0$, for almost everywhere[4] all $\omega \in \Omega$, there exists two random sequences $\{t_n = t_n(\omega)\}_{n \in \mathbb{N}}$ and $\{s_n = s_n(\omega)\}_{n \in \mathbb{N}}$, both tending to infinity, such that:

$$B(t_n, \omega) \;\geq\; (1 - \varepsilon)\sqrt{2t_n \ln|\ln(t_n)|} \quad \text{and} \quad B(s_n, \omega) \;\leq\; -(1 - \varepsilon)\sqrt{2s_n \ln|\ln(s_n)|}\,.$$

[4] A property holds "almost everywhere" (a.e.) if the set of elements for which the property does not hold is a set of measure zero.

Since the map $t \mapsto B(t,\omega)$ is a.s. continuous, these last two inequalities prove that, for $\omega \in \Omega$ a.e., $B(t,\omega)$ visits every real number infinitely many times (more precisely, it changes sign infinitely many times) in every neighbourhood $[M, +\infty)$ of infinity.

Using the properties of invariance of SBM, it is possible to extend the Theorem 1.1.6, which is about asymptotic properties, to local ones on paths behaviour.

Theorem 1.1.7 (**Local law of the iterate logarithm**) *For all given $t_0 \geq 0$, if $\{B(t)\}_{t\in[0,+\infty)}$ is a standard Brownian motion, the following limits hold almost surely:*

$$\limsup_{h\to 0} \frac{B(t_0+h)-B(t_0)}{\sqrt{2\cdot h\cdot \ln\left|\ln\left(\frac{1}{h}\right)\right|}} = 1 \quad \textit{and} \quad \liminf_{h\to 0} \frac{B(t_0+h)-B(t_0)}{\sqrt{2\cdot h\cdot \ln\left|\ln\left(\frac{1}{h}\right)\right|}} = -1\,. \tag{1.6}$$

As shown for the previous theorem, this limits are able to describe the local paths movement. For all $\varepsilon > 0$, for $\omega \in \Omega$ a.e., there exists two random sequences $\{h_n = h_n(\omega)\}_{n\in\mathbb{N}}$ and $\{k_n = k_n(\omega)\}_{n\in\mathbb{N}}$, both tending to zero, such that:

$$B(h_n,\omega) \geq (1-\varepsilon)\sqrt{2h_n \ln\left|\ln\left(\frac{1}{h_n}\right)\right|} \quad \text{and} \quad B(k_n,\omega) \leq -(1-\varepsilon)\sqrt{2k_n \ln\left|\ln\left(\frac{1}{k_n}\right)\right|}\,.$$

Because of the continuity of paths, we have that $B(t,\omega)$ changes sign infinitely many times in every right neighbourhood $[t_0, \delta)$. This behaviour of SBM proves heuristically the following result.

Corollary 1.1.8 *Almost every path of the standard Brownian motion has infinite variation on any interval, no matter how small it is.*

Next, we show the non-differentiability of its paths.

Theorem 1.1.9 *Let $\{B(t)\}_{t\in[0,+\infty)}$ be a standard Brownian motion. For every time $t_0 \in [0,+\infty)$, $B(t)$ is not differentiable in $t = t_0$ almost surely.*

Proof: Theorem 1.1.7 ensures that the following limits are valid almost surely:

$$\limsup_{h\to 0} \frac{B(t_0+h)-B(t_0)}{h} = \limsup_{h\to 0} \frac{\sqrt{2\cdot h\cdot \ln\left|\ln\left(\frac{1}{h}\right)\right|}}{h} = \\ = \limsup_{h\to 0} \sqrt{\frac{2\cdot \ln\left|\ln\left(\frac{1}{h}\right)\right|}{h}} = +\infty$$

and

$$\liminf_{h\to 0} \frac{B(t_0+h)-B(t_0)}{h} = \liminf_{h\to 0} -\frac{\sqrt{2\cdot h\cdot \ln\left|\ln\left(\frac{1}{h}\right)\right|}}{h} = \\ = \liminf_{h\to 0} -\sqrt{\frac{2\cdot \ln\left|\ln\left(\frac{1}{h}\right)\right|}{h}} = -\infty\,.$$

Hence, almost surely, we have that:

$$\nexists\, B'(t_0) = \lim_{h \to 0} \frac{B(t_0 + h) - B(t_0)}{h}\,. \tag{1.7}$$

Hence every real standard Brownian motion is a path-continuous nowhere differentiable stochastic process.

□

Now we will produce an alternative definition of SBM, which involves concepts such as *filtration* and *process adapted to a filtration*. Of course, this definition is equivalent to the one given at the beginning of this section.

Proposition 1.1.10 *A real stochastic process $\{B(t)\}_{t \in T}$ is a standard Brownian motion if and if only it satisfies the properties* **(a)**, **(c)** *and* **(d)** *of* Definition 1.1.1 *and if the following one holds:*

(b') *for all $0 \leq s < t$, the random variable $B(t) - B(s)$ is independent from the σ-algebra $\mathcal{F}_s^B = \sigma\left(\{B(u)\}_{u \in [0,s] \cap T}\right)$.*

We will not report the proof of this result since it is quite trivial. However, we want to remark that the proposition is only a special case of a more general situation: *Any* stochastic process $\{X(t)\}_{t \in [0,+\infty)}$ has independent increments, so **(b)** of Definition 1.1.1 holds, if and if only the previous condition **(b')** is valid.

The following theorem proves that a SBM is a even a martingale.

Theorem 1.1.11 *The standard Brownian motion $\{B(t)\}_{t \in T}$, defined on a filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t \in T}, \mathbf{P})$ and $\mathcal{F}_t$ – adapted, is a $\{\mathcal{F}_t\}_{t \in T}$ – martingale.*

Proof: Given $0 \leq s < t$, it follows almost surely:

$$\mathbb{E}[B(t)|\,\mathcal{F}_s] = \mathbb{E}[B(t) - B(s) + B(s)|\,\mathcal{F}_s] = \mathbb{E}[B(t) - B(s)|\,\mathcal{F}_s] + \mathbb{E}[B(s)|\,\mathcal{F}_s]\,.$$

Due to Proposition 1.1.10, the increments are independent on the natural filtration $\mathcal{F}_s^B$ and hence on the filtration $\mathcal{F}_s$, since we have $\mathcal{F}_s^B \subseteq \mathcal{F}_s$. Therefore, the standard Brownian motion is a $\{\mathcal{F}_t\}_{t \in T}$ – martingale:

$$\mathbb{E}[B(t)|\,\mathcal{F}_s] = \mathbb{E}[B(t) - B(s)|\,\mathcal{F}_s] + \mathbb{E}[B(s)|\,\mathcal{F}_s] = \mathbb{E}[B(s)|\,\mathcal{F}_s] = B(s)\,.$$

□

The last theorem of this section is about self-affinity. As a matter of fact, SBM is a H–s.a. process with Hurst exponent equal to $\frac{1}{2}$.

Theorem 1.1.12 *The standard Brownian motion* $\{B(t)\}_{t\in[0,+\infty)}$ *is a self-affine stochastic process with* $H = \frac{1}{2}$.

Proof: By virtue of Theorem 1.1.3, we have:

$$\begin{aligned}\Pr\{B(t) \le b\} &= \frac{1}{\sqrt{2\cdot\pi\cdot\mathbb{V}[B(t)]}}\int_{-\infty}^{b} e^{-\frac{\{x-\mathbb{E}[B(t)]\}^2}{2\cdot\mathbb{V}[B(t)]}}\,\mathrm{d}x \\ &= \frac{1}{\sqrt{2\cdot\pi\cdot t}}\int_{-\infty}^{b} e^{-\frac{x^2}{2\cdot t}}\,\mathrm{d}x\,.\end{aligned}$$

To prove that $\{B(t)\}$ is H–s.a. we calculate the probability of the rescale-timed process with $a > 0$, that is:

$$\Pr\{B(a\cdot t) \le b\} = \frac{1}{\sqrt{2\cdot\pi\cdot(a\cdot t)}}\int_{-\infty}^{b} e^{-\frac{x^2}{2\cdot(a\cdot t)}}\,\mathrm{d}x\,.$$

Moreover, since $B(t)$ is normal distributed with zero mean and variance equal to t, the same variable multiplied by $a^{\frac{1}{2}}$ is normally distributed as well, with mean and variance:

$$\sqrt{a}\cdot B(t) \overset{d}{\sim} \mathcal{N}(0, a\cdot t)\,. \tag{1.8}$$

The probability that variable is less than a given number is therefore:

$$\Pr\left\{\sqrt{a}\cdot B(t) \le b\right\} = \frac{1}{\sqrt{2\cdot\pi\cdot(a\cdot t)}}\int_{-\infty}^{b} e^{-\frac{x^2}{2\cdot(a\cdot t)}}\,\mathrm{d}x\,.$$

Since $\Pr\{\sqrt{a}\cdot B(t) \le b\} = \Pr\{B(a\cdot t) \le b\}$, we have

$$\left\{B(a\cdot t)\right\}_{t\ge 0} \overset{d}{\sim} \left\{a^{\frac{1}{2}}\cdot B(t)\right\}_{t\ge 0}$$

that gives the thesis. Hence the standard Brownian motion is a self-affine process with Hurst exponent equal to $H = \frac{1}{2}$.

□

We have already proved that SBM has independent and stationary increments. Together with the previous result, it turns out to be a H–s.a.s.i. process as well. As a matter of fact, we have:

1. $B(0) = 0$ a.s., from Definition 1.1.1;
2. $\mu_B(t) = 0$, from Theorem 1.1.3;
3. $\{B(-t)\}_{t\ge 0} \overset{d}{\sim} \{-B(t)\}_{t\ge 0}$, from Proposition 1.1.4;
4. $\sigma_B^2(t) = t$ from Theorem 1.1.3 and 1.1.12[5];

[5]Since the Brownian motion is Standard, we have $\sigma_B^2(1) = \sigma^2 = 1$. In addiction, if considered only positive times, the expression of the process variance becomes $\sigma_B^2(t) = t^{2H}$. Eventually, because the Hurst exponent is equal to $\frac{1}{2}$, it follows $\sigma_B^2(t) = t$.

5. $\mathcal{C}_B(t,s) = \min\{t,s\}$, from Theorem 1.1.3 and 1.1.12[6];

6. $H = \frac{1}{2} \in (0,1)$, from Theorem 1.1.12.

That is, the standard Brownian motion is a self-affine stochastic process with stationary increments and Hurst exponent $H = \frac{1}{2}$.
Eventually we introduce (non-standard) Brownian motions (BM), named Arithmetical Brownian Motion[7]as well.

Definition 1.1.13 *Given a probability space* $(\Omega, \mathcal{F}, \mathbf{P})$*, every real stochastic process* $\{B^*(t)\}_{t\in[0,+\infty)}$ *defined on it is said* Brownian motion*, if it satisfies the properties of the* Definition 1.1.1*, except for the letter* **(c)** *which is replaced by a more general one:*

(c') *for every* $0 \leq s < t$*, all increments* $B^*(t) - B^*(s)$ *are Gaussian with mean and variance proportional with the width of the time interval, that is:*

$$B^*(t) - B^*(s) \overset{d}{\sim} \mathcal{N}[\mu \cdot (t-s), \sigma^2 \cdot (t-s)] \tag{1.9}$$

where $\mu = \mu_{B^*}(1)$ *and* $\sigma^2 = \sigma^2_{B^*}(1)$.

Often μ and σ are respectively called *drift* and *diffusion parameter*. Most of the previous results on SBM can be properly extended to it, such as

$$\mathbb{E}[B^*(t)] = \mu \cdot t \qquad \mathbb{V}[B^*(t)] = \sigma^2 \cdot t$$

and the non-differentiability of its paths.

As it is well-known, a normal random variable can be obtained as linear transformation of the standard normal one. At the same way, a Brownian motion $\{B^*(t)\}$ can be represented as linear transformation of a standard Brownian motion $\{B(t)\}$. Indeed:

$$B^*(t) = \mu \cdot t + \sigma \cdot B(t)\,. \tag{1.10}$$

1.2 Stochastic integration and Gaussian white noise

This section is intended to give only few instruments about stochastic integration, without claiming to be exhaustive. We shall produce only the necessary tools for a correct understanding of the following subjects. For a more in-depth approach, see [19], [12] and [34].

[6]Considering only positive times such that $0 \leq s < t$, the covariance of the process becomes $\mathcal{C}_B(t,s) = \frac{1}{2} \cdot \left[t^{2H} + s^{2H} + (t-s)^{2H}\right]$. Eventually, because the Hurst exponent is equal to $\frac{1}{2}$, it follows $\mathcal{C}_B(t,s) = \frac{1}{2} \cdot (t+s-t+s) = s = \min\{t,s\}$.

[7]The adjective "arithmetical" is used to underline the fact that the process increments are independent, unlike in the Geometric Brownian Motion case in which the independence concerns the logarithms of the process ratios (returns).

Given a integrand variable $X(t)$ which can be, for the moment, both a deterministic function or a stochastic process and $\{B(t)\}_{t\in T}$ which is a standard Brownian motion, we are now interested in understanding on what conditions the following integral

$$\int_a^b X(t)\,\mathrm{d}B(t) \tag{1.11}$$

makes sense. The first problem we have to focus on is about the variable of integration that is a standard Brownian motion. As a matter of fact, we have seen that Corollary 1.1.8 and Theorem 1.1.9 state that SBM has infinite variations on any interval and is nowhere differentiable: Hence traditional integration techniques do not consistently allow the convergence of the integral. Firstly, let we consider the case when the integrand is a deterministic function, that is depending only on $t \in T$ and not on $\omega \in \Omega$ (however, the dependence on ω will always be omitted in the formulas):

$$\int_a^b f(t)\,\mathrm{d}B(t)$$

which is called *Wiener integral*. We want to underline a first "approximation": It is true that the SBM is not differentiable, but its instantaneous increments can be seen as the limit for $\Delta t \to 0$ of the increment $B(t+\Delta t)-B(t)$. In this way, the Weiner integral becomes a normal random variable with null mean and variance equal to $\int_a^b f(t)^2\,\mathrm{d}t$.

However, Wiener integral is rigorously defined if and if only (this criterion is valid for whatever stochastic process as integrator):

- the stochastic process has uncorrelated increments (that is for SBM, being even independent);
- the function $f(t)$ is a Lebesgue square-integrable functions.

As a matter of fact, it is possible to define a integral $\int_a^b \varphi(t)\,\mathrm{d}B(t)$, where $\varphi(t)$ is a staircase functions. Then we consider a sequence of staircase function $\varphi_n(t)$ that approximates $f(t)$ in mean square[8]. Finally, the Wiener integral is defined as the limit in mean square[9] of the sequences of random

[8] According to the *Labesgue's dominated convergence theorem*, let $A \subset \mathbb{R}^n$ be a measurable set and consider a sequence of measurable functions $h_n : A \to \mathbb{R}$, such that $\exists h : A \to \mathbb{R}$ measurable with $h_n(\mathbf{x}) \to h(\mathbf{x})$ a.e. for $\mathbf{x} \in A$ when $n \to \infty$. If $\exists g : \int_A |g|^2\,\mathrm{d}\mu < +\infty$ such that $|h_n(\mathbf{x})| \le g(\mathbf{x})$ a.e. for $\mathbf{x} \in A$ and $\forall n \in \mathbb{N}$, hence $h_n \to h$ in mean square.

[9] We say that the sequence of random variables X_n converges in mean square towards the random variable X, if $\mathbb{E}(|X_n|^2)$ and $\mathbb{E}(|X|^2)$ exist and are finite, and $\lim_{n\to\infty} \mathbb{E}\left(|X_n - X|^2\right) = 0$. To describe concisely that property, it is usually written:

$$X_n \xrightarrow{L^2} X \qquad \text{or} \qquad \lim_{n\to\infty} \text{m.s.}\, X_n = X\,.$$

variables defined by $\int_a^b \varphi_n(t)\ \mathrm{d}B(t)$. This procedure leads to a completely coherent definition; in addiction, if $f(t)$ is continuous, the integral can be calculated with a different method similar to the one used to introduce the Riemann integral, based on the partition of the interval of integration $[a, b]$.

The next step to face and solve, if possible, is how compute the more general integral

$$\int_a^b X(t)\ \mathrm{d}B(t)$$

where the integrand is a stochastic process itself. A rigorous explanation of it is due to the Japanese mathematician Kiyoshi Itō (1944). For this reason, that random variable is usually called *Itō integral*. The definition of such a stochastic integral requires the integrand process to satisfy some rather general conditions, whose most important is the *non-anticipativity* with respect the standard Brownian motion.

Definition 1.2.1 *Given a standard Brownian motion* $\{B(t)\}_{t\in T}$ *and its natural filtration* $\mathcal{F}_t^B = \sigma(B(s), s \leq t)$, *we say that a stochastic process* $\{X(t)\}_{t\in T}$ *is* non-anticipativity *with respect the SBM if:*

- $\{X(t)\}$ *is* $\mathcal{F}_t$*–measurable;*
- $X(t)$ *is independent on the increments* $B(\tau) - B(t)$, *with* $\tau > t$.

That is, $\{X(t)\}$ *depend on the past and the present of* $\{B(t)\}$, *but not on its future.*

This concept, that expresses a natural requirement of causality, is fundamental to correctly define Itō integral. In particular it is possible to prove that the following processes are non-anticipative:

$$B(t) \qquad \int_{t_0}^t \vartheta\left[B(h)\right]\ \mathrm{d}h \qquad \int_{t_0}^t \vartheta\left[B(h)\right]\ \mathrm{d}B(h)\,.$$

Moreover, this property is valid even for

$$\int_{t_0}^t X(h)\ \mathrm{d}h \qquad \int_{t_0}^t X(h)\ \mathrm{d}B(h)$$

if $\{X(h)\}$ is non-anticipative.

Below we will show explicitly what happens if we try to integrate a SBM with respect itself. All the cautions involved and how to define this special Itō integral will be exhibited. Given

$$\int_{t_0}^{t_n} B(t)\ \mathrm{d}B(t)$$

we could be tempted to define it as done for the Riemann integral with a limit of summations. Due to summation by parts, surely it hold[10]

$$\underbrace{\sum_{i=0}^{n-1} B(t_i)\cdot[B(t_{i+1})-B(t_i)]}_{S_1} = B(t_n)^2 - B(t_0)^2 - \underbrace{\sum_{i=0}^{n-1} B(t_{i+1})\cdot[B(t_{i+1})-B(t_i)]}_{S_2}\,.$$

If we could define the Itō integral as the Riemann one, we would expect that S_1 and S_2 converge to the same limit, but in fact, this is not true. We are going to show it.

First of all, we see that

$$\begin{aligned} S_2 - S_1 &= \sum_{i=0}^{n-1} B(t_{i+1})\cdot[B(t_{i+1})-B(t_i)] - \sum_{i=0}^{n-1} B(t_i)\cdot[B(t_{i+1})-B(t_i)] = \\ &= \sum_{i=0}^{n-1}[B(t_{i+1})-B(t_i)]\cdot[B(t_{i+1})-B(t_i)] = \sum_{i=0}^{n-1}[B(t_{i+1})-B(t_i)]^2 \end{aligned}$$

and

$$S_2 + S_1 = B(t_n)^2 - B(t_0)^2$$

directly from the first expression. Hence we have:

$$\begin{aligned} S_1 &= \frac{B(t_n)^2 - B(t_0)^2}{2} - \frac{\sum_{i=0}^{n-1}[B(t_{i+1})-B(t_i)]^2}{2} \\ S_2 &= \frac{B(t_n)^2 - B(t_0)^2}{2} + \frac{\sum_{i=0}^{n-1}[B(t_{i+1})-B(t_i)]^2}{2}\,. \end{aligned}$$

Given the partition $\Pi = \{t_0, t_1, ..., t_{n-1}, t_n\}$ of the time interval $[t_0, t_n]$, we can fix $t_0 = s$ and $t_n = t$ without loss of generality. Thus, we have:

$$\begin{aligned} S_1 &= \frac{B(t)^2 - B(s)^2}{2} - \frac{\sum_{i=0}^{n-1}[B(t_{i+1})-B(t_i)]^2}{2} \\ S_2 &= \frac{B(t)^2 - B(s)^2}{2} + \frac{\sum_{i=0}^{n-1}[B(t_{i+1})-B(t_i)]^2}{2}\,. \end{aligned}$$

The last step is to study the convergence of the sum $\sum_{i=0}^{n-1}[B(t_{i+1})-B(t_i)]^2$.

Proposition 1.2.2 *Given a standard Brownian motion $\{B(t)\}_{t\geq 0}$, the sum $\sum_{i=0}^{n-1}[B(t_{i+1})-B(t_i)]^2$ converges in mean square towards the length of the time interval $t-s$.*

[10] Given two function $f(x)$ and $g(x)$, this formula state that for $x \in [x_0, x_n]$, we have:

$$\sum_{i=0}^{n-1} g(x_i)\cdot[f(x_{i+1})-f(x_i)] = f(x_n)\cdot g(x_n) - f(x_0)\cdot g(x_0) - \sum_{i=0}^{n-1} f(x_{i+1})\cdot[g(x_{i+1})-g(x_i)]$$

Proof: From the properties of SBM we have $B(t_{i+1}) - B(t_i) \overset{d}{\sim} \mathcal{N}(0, t_{i+1} - t_i)$, so:

$$\begin{aligned} B(t_{i+1}) - B(t_i) &= \frac{\sqrt{t_{i+1} - t_i}}{\sqrt{t_{i+1} - t_i}} \cdot [B(t_{i+1}) - B(t_i)] = \\ &= \sqrt{t_{i+1} - t_i} \cdot \frac{B(t_{i+1}) - B(t_i)}{\sqrt{t_{i+1} - t_i}} = \\ &= \sqrt{t_{i+1} - t_i} \cdot \varepsilon \end{aligned}$$

where $\varepsilon \overset{d}{\sim} \mathcal{N}(0,1)$, since that quantity does not depend on t_i. Substituting the last expression in the sum and taking the expected value, we obtain[11] :

$$\begin{aligned} \mathbb{E}\left\{\sum_{i=0}^{n-1}[B(t_{i+1}) - B(t_i)]^2\right\} &= \mathbb{E}\left[\sum_{i=0}^{n-1}(\sqrt{t_{i+1} - t_i} \cdot \varepsilon)^2\right] = \mathbb{E}\left[\sum_{i=0}^{n-1}(t_{i+1} - t_i) \cdot \varepsilon^2\right] = \\ &= \sum_{i=0}^{n-1}(t_{i+1} - t_i) \cdot \mathbb{E}(\varepsilon^2) = \sum_{i=0}^{n-1} t_{i+1} - t_i = \\ &= (t_1 - t_0) + (t_2 - t_1) + ... + (t_n - t_{n-1}) = \\ &= t_n - t_0 = t - s\,. \end{aligned}$$

The last step to prove the convergence in mean square is that the variance of the sum tends to zero, that is:

$$\begin{aligned} \mathbb{V}\left\{\sum_{i=0}^{n-1}[B(t_{i+1}) - B(t_i)]^2\right\} &= \mathbb{E}\left\{\left\{\sum_{i=0}^{n-1}[B(t_{i+1}) - B(t_i)]^2 - \mathbb{E}\left\{\sum_{i=0}^{n-1}[B(t_{i+1}) - B(t_i)]^2\right\}\right\}^2\right\} = \\ &= \mathbb{E}\left\{\left[\sum_{i=0}^{n-1}(t_{i+1} - t_i) \cdot \varepsilon^2 - \sum_{i=0}^{n-1} t_{i+1} - t_i\right]^2\right\} = \\ &= \mathbb{E}\left\{\left[\sum_{i=0}^{n-1}(t_{i+1} - t_i) \cdot (\varepsilon^2 - 1)\right]^2\right\} = \\ &= \mathbb{E}\left[(\varepsilon^2 - 1)^2 \cdot \left(\sum_{i=0}^{n-1} t_{i+1} - t_i\right)^2\right] = \\ &= \mathbb{E}\left[(\varepsilon^2 - 1)^2\right] \cdot (t_n - t_0)^2 = \\ &= \mathbb{E}(\varepsilon^4 - 2 \cdot \varepsilon^2 + 1) \cdot (t - s)^2\,. \end{aligned}$$

Since $\mathbb{E}(\varepsilon^4) = 2$ and $\mathbb{E}(\varepsilon^2) = 1$, we have:

$$0 \leq \mathbb{V}\left\{\sum_{i=0}^{n-1}[B(t_{i+1}) - B(t_i)]^2\right\} = (t - s)^2 \leq \max_{0 \leq i \leq n-1} |t_{i+1} - t_i| \cdot (t - s)\,.$$

Since the norm of the partition $\Pi = \{s = t_0, t_1, ..., t_{n-1}, t_n = t\}$ is the length of the longest of its subintervals, that is $\pi = \max_{0 \leq i \leq n-1} |t_{i+1} - t_i|$, when

[11] If $\varepsilon \overset{d}{\sim} \mathcal{N}(0,1)$, it follows that ε^2 is a chi-square with one degree of freedom. Hence, $\mathbb{E}(\varepsilon^2) = 1$ and $\mathbb{V}(\varepsilon^2) = \mathbb{E}(\varepsilon^4) = 2$.

$\pi \to 0$ (π becomes dense in $[s,t]$) by the squeeze theorem it follows that:

$$\lim_{\pi\to 0} \mathbb{V}\left\{\sum_{i=0}^{n-1}[B(t_{i+1}) - B(t_i)]^2\right\} = 0\,.$$

Hence the sum of squared increments converges to its mean:

$$\lim_{\pi\to 0}\sum_{i=0}^{n-1}[B(t_{i+1}) - B(t_i)]^2 = \mathbb{E}\left\{\sum_{i=0}^{n-1}[B(t_{i+1}) - B(t_i)]^2\right\} = t - s\,.$$

It is the equivalent to saying

$$\sum_{i=0}^{n-1}[B(t_{i+1}) - B(t_i)]^2 \overset{L^2}{\to} t - s$$

that is the thesis.

□

This result allows us to write:

$$S_1 = \sum_{i=0}^{n-1} B(t_i)\cdot[B(t_{i+1}) - B(t_i)] \overset{L^2}{\longrightarrow} \frac{B(t)^2 - B(s)^2}{2} - \frac{t-s}{2}$$

$$S_2 = \sum_{i=0}^{n-1} B(t_{i+1})\cdot[B(t_{i+1}) - B(t_i)] \overset{L^2}{\longrightarrow} \frac{B(t)^2 - B(s)^2}{2} + \frac{t-s}{2}$$

from which it follows that the Riemann integral of $B(t)$ with respect itself *does not exist*, since the limit of $\sum_{i=0}^{n-1} B(\tau_i)\cdot[B(t_{i+1}) - B(t_i)]$ depends on the choice of $\tau_i \in [t_i, t_{i+1}]$. Hence, at this point we should choose a "proper" τ_i on which calculate the integral $\int_s^t B(h)\,\mathrm{d}B(h)$. If we opt for the left endpoint of the interval $\tau_i = t_i$, we get the Itō integral[12]:

$$\int_s^t B(h)\,\mathrm{d}B(h) = \frac{B(t)^2 - B(s)^2}{2} - \frac{t-s}{2}\,. \tag{1.12}$$

This definition of stochastic integral implies a different set of rules about integration and differentiation. Since the Riemann integral does not exist, the

[12] At first glance, it may seem more appropriate to select the midpoint of the interval $\tau_i = \frac{t_{i+1}-t_i}{2}$. This choice, deeply inspired by the Riemann procedure, leads to the definition of Stratonovich integral, which correspond to:

$$\int_s^t B(h)\,\mathrm{d}B(h) = \frac{B(t)^2 - B(s)^2}{2}\,.$$

Hence the rules of stochastic integration are the same of Riemann's ones. So why to choose Itō integral? Because it does own few attractive properties such as being a martingale. Moreover, it has not been possible to prove yet the convergence and coherence of Stratonovich integral to Itō's one.

classical formulas are not valid anymore (in particular the total differential rule needs to be changed).

Now below, we shall only state few properties of stochastic integration.

Proposition 1.2.3 *Given to real valued stochastic processes* $\{X(t)\}_{t\in T}$ *and* $\{Y(t)\}_{t\in T}$ *and a standard Brownian motion, the following relations are valid:*

$$\mathbb{E}\left[\int_a^b X(t)\,\mathrm{d}B(t)\right] = 0$$

$$\mathbb{V}\left[\int_a^b X(t)\,\mathrm{d}B(t)\right] = \mathbb{E}\left[\int_a^b X(t)^2\,\mathrm{d}t\right]$$

$$\mathbb{C}\mathrm{ov}\left[\int_a^b X(t)\,\mathrm{d}B(t), \int_a^b Y(t)\,dB(t)\right] = \mathbb{E}\left[\int_a^b X(t)\cdot Y(t)\,\mathrm{d}t\right]$$

$$\mathbb{E}\left[\int_a^b X(t)\,\mathrm{d}B(t)\right] = \mathbb{E}\left[\int_a^b Y(t)\,\mathrm{d}B(t)\right] \text{ a.s. } \Leftrightarrow \mathbb{E}\left\{\int_a^b [X(t)-Y(t)]^2\,\mathrm{d}t\right\} = 0\,.$$

The last topic to introduce is the *Gaussian white noise* process. From a strictly formal point of view, the Gaussian white noise does not exist as a stochastic process such as the Dirac delta function does not exist as a function[13]. Let we proceed with a definition.

Definition 1.2.4 *Every real Gaussian stochastic process* $\{W(t)\}_{t\in[0,+\infty)}$ *with null mean and autocovariance function equal to* $\mathcal{C}_W(t,s) = \delta(t-s)$ *is said* Gaussian white noise.

The fact that such a function does not exist, shows that the definition of $\{W(t)\}_{t\in[0,+\infty)}$ as a real Gaussian process is ill posed. However, we want to show that the stochastic integral

$$B(t) := \int_0^t W(h)\,\mathrm{d}h \tag{1.13}$$

is a standard Brownian motion. Since $\{W(t)\}$ is a Gaussian process with null mean, also $\{B(t)\}$ is normal with zero mean, being a linear function of

[13]The Dirac delta function is the "function" such that the following Lebesgue integrals hold

$$\int_A \delta(x)\,\mathrm{d}x = 1 \qquad \int_{\neg A} \delta(x)\,\mathrm{d}x = 0$$

where $A \subset \mathbb{R} : \{0\} \in A$.

$W(t)$. Let we calculate its autocovariance function:

$$\begin{aligned}\mathcal{C}_B(t,s) &= \mathbb{E}\left[\left(\int_0^s W(h)\,\mathrm{d}h\right)\cdot\left(\int_0^t W(k)\,\mathrm{d}k\right)\right] = \\ &= \int_0^s\int_0^t \mathbb{E}[W(h)\cdot W(k)]\,\mathrm{d}h\,\mathrm{d}k = \int_0^s\int_0^t \delta(k-h)\,\mathrm{d}h\,\mathrm{d}k = \\ &= \int_0^s \mathbf{1}_{[0,t)}(k)\,\mathrm{d}k = \min\{s,t\}\,.\end{aligned}$$

Due to Theorem 1.1.3, $\{B(t)\}_{t\in[0,+\infty)}$ is a SMB.

From the definition $B(t) = \int_0^t W(h)\,\mathrm{d}h$, it follows that the Gaussian white noise can be *thought* as the first derivative of a standard Brownian motion that, we know, it does not exist. For this reason, it is sometimes called *formal derivative*. Thus we can write[14]

$$W(t) = \frac{\mathrm{d}B(t)}{\mathrm{d}t} \quad \text{or} \quad \mathrm{d}B(t) = W(t)\,\mathrm{d}t \tag{1.14}$$

being the second one more advisable.

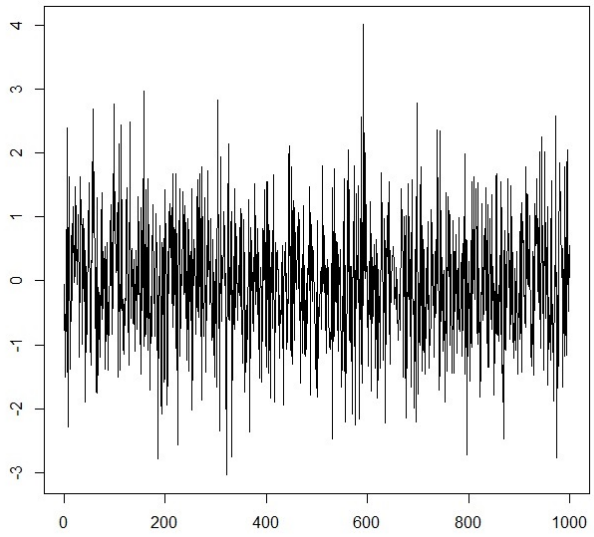


Figure 1.2: Discrete version of a Gaussian white noise

[14]This fact can be seen as the same thing of saying that the Dirac delta function is the derivative of the Heaviside step function

$$H(x) = \begin{cases}1 & \text{if } x \geq 0\\ 0 & \text{if } x < 0\end{cases} = \int_{-\infty}^{x} \delta(s)\,\mathrm{d}s \quad \Longleftrightarrow \quad \mathrm{d}H(x) = \delta(x)\,\mathrm{d}x\,.$$

1.3 Fractional Brownian Motion (FBM)

We have seen that one of the most important stochastic processes used in a variety of applications is the standard Brownian motion which is a Gaussian process with zero mean and covariance function $\min\{s,t\}$. That process has independent increments and its formal derivative, the white noise, is used as input noise in dynamical systems, giving rise to stochastic differential equations.

Motivated from some applications in hydrology, telecommunications, theory of queues and mathematical finance, there has been a new request in input noises *without* independent increments and possessing long-range dependence and self-affine properties. Long-range dependence in a stationary time series occurs when the covariances tend to zero like a power function and so slowly that their sums diverge. We have already known that self-affinity property means invariance in distribution under a suitable change of scale. One of the simplest stochastic processes which is Gaussian, self-affine and it has stationary increments is fractional Brownian motion (FBM), which is a generalization of the classical Brownian motion. As we shall see later, the fractional Brownian motion possesses long-range dependence when its Hurst exponent is larger than $\frac{1}{2}$.

Below we survey some properties of the fractional Brownian motion, and describe different representations of it.

Definition 1.3.1 *Given a filtered probability space* $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\in[0,+\infty)}, \mathbf{P})$, *every real stochastic process* $\{B_H(t)\}_{t\in[0,+\infty)}$ *defined on it is said* fractional Brownian motion, *if it satisfies the following properties:*

(a) *it is a Gaussian process;*

(b) *it is* $\mathcal{F}_t$ *– adapted;*

(c) *it is a* H*–s.a.s.i. stochastic process.*

From the definition of H–s.a.s.i. process (see Appendix A), it follows that the process mean is null, so

$$\mu_{B_H}(t) = 0 \tag{1.15}$$

and the autocovariance function, for $s,t \geq 0$, is given by:

$$\mathcal{C}_{B_H}(t,s) = \frac{\sigma^2}{2}\cdot\left(t^{2H} + s^{2H} - |t-s|^{2H}\right). \tag{1.16}$$

Being the process normal, it is fully described by this set of information. From the definition, the distribution of the increments is normal with mean

$$\mathbb{E}[B_H(t) - B_H(s)] = \mu_{B_H}(t) - \mu_{B_H}(s) = 0$$

and variance

$$\begin{aligned}\mathbb{E}\left\{[B_H(t) - B_H(s)]^2\right\} &= \mathbb{V}[B_H(t) - B_H(s)] = \\ &= \mathbb{V}[B_H(t)] + \mathbb{V}[B_H(s)] - 2 \cdot \mathbb{E}[B_H(t) \cdot B_H(s)] = \\ &= \mathcal{C}_{B_H}(t,t) + \mathcal{C}_{B_H}(s,s) - 2 \cdot \mathcal{C}_{B_H}(t,s) = \\ &= \sigma^2 \cdot t^{2H} + \sigma^2 \cdot s^{2H} - \sigma^2 \cdot \left(t^{2H} + s^{2H} - |t-s|^{2H}\right) = \\ &= \sigma^2 \cdot |t-s|^{2H} .\end{aligned}$$

This shows the stationarity of the increments, since the second moment depends only on the distance between s and t, not on the time instants themselves. In addiction, it is possible to prove that:

$$\mathbb{E}\left[|B_H(t) - B_H(s)|^k\right] = \mathbb{E}\left[|B_H(1)|^k\right] \cdot |t-s|^{kH} . \tag{1.17}$$

for every $k \geq 1$. Of course, if we take $0 \leq s < t$, the absolute value in the right side can be removed. Using the following theorem, this last result implies the following basic feature of FBM. For its proof see Section 3.4.

Theorem 1.3.2 (Kolmogorov's continuity theorem) *Let $\{X(t)\}_{t \in \mathbb{R}_0^+}$ be a stochastic process. If for all times t, there exist positive constants α, β and K such that*

$$\mathbb{E}\big[|X(t) - X(s)|^\alpha\big] \leq K \cdot |t-s|^{1+\beta}$$

thus $\{X_t\}$ can be modified for each t on a set of measure zero, to obtain an equivalent version that is almost surely continuous.

This theorem allows us to state the following.

Proposition 1.3.3 *The fractional Brownian motion has continuous paths almost surely.*

Proof: Due to Kolmogorov's continuity theorem and to the property of n-th moment of the FBM increments, we have $\alpha = k$, $\beta = \alpha H - 1$ and $K = \mathbb{E}\left[|B_H(1)|^k\right]$.

□

As in the SBM case, it is possible to prove that a fractional Brownian motion is nowhere differentiable (but we will not prove it explicitly). As a matter of fact, Marcus [29] showed:

$$\limsup_{h \to 0} \sup_{0 \leq t \leq 1} \frac{B_H(t_0 + h) - B_H(t_0)}{h^H \cdot \sqrt{2 \cdot \ln\left(\frac{1}{h}\right)}} = 1 \qquad \text{a.s.} \tag{1.18}$$

leading to a infinite derivative, that is, the non existence of it.

As in the SBM case (see Theorem 1.1.3), we can introduce an alternative characterization of FBM.

Proposition 1.3.4 *Let $\{B_H(t)\}_{t\in[0,+\infty)}$ be a real stochastic process defined on a filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\in[0,+\infty)}, \mathbf{P})$. If the following properties hold:*

(a) $B_H(0) = 0$ *almost surely;*

(b) $\{B_H(t)\}$ *is $\mathcal{F}_t$ – adapted;*

(c) *for all $t \geq 0$*

$$B_H(t) \overset{d}{\sim} \mathcal{N}\left(0, \sigma^2 \cdot t^{2H}\right) \tag{1.19}$$

with $\sigma > 0$ and $0 < H < 1$;

(d) *the increments are stationary.*

Hence the stochastic process $\{B_H(t)\}_{t\in[0,+\infty)}$ is a fractional Brownian motion.

The equivalence between the two different statements is trivial.

As previously said, the process is said a *standard* fractional Brownian motion if ${\sigma^2}_{B_H}(t) = \sigma^2 = 1$, thus implying:

$$\mathcal{C}_{B_H}(t,s) = \frac{1}{2} \cdot \left(t^{2H} + s^{2H} - |t-s|^{2H}\right).$$

If not explicitly said, from now on with FBM we will always refer to the standard fractional Brownian motion.

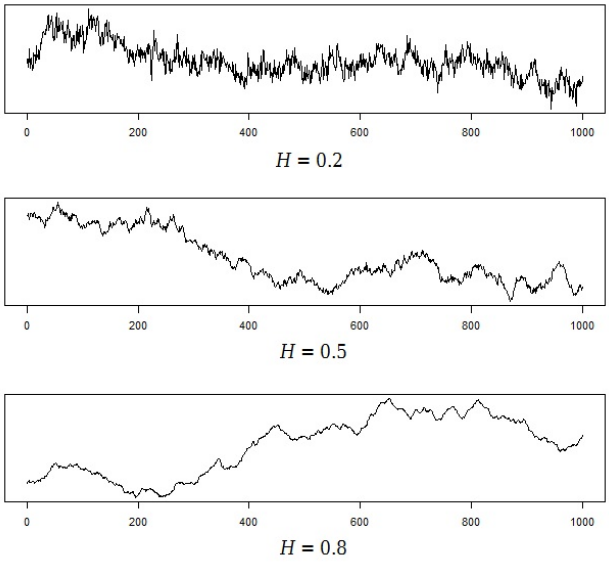


Figure 1.3: Discrete versions of standard fractional Brownian motions

Note that, if $H = \frac{1}{2}$, the FBM is nothing but a SMB. As a matter of fact the autocovariance function becomes (the mean is not affected by H):

$$\mathcal{C}_{B_{\frac{1}{2}}}(t,s) = \frac{1}{2} \cdot \left(t^{2\cdot\frac{1}{2}} + s^{2\cdot\frac{1}{2}} - |t-s|^{2\cdot\frac{1}{2}} \right) = \frac{1}{2} \cdot (t + s - |t-s|) = \min\{s,t\}\,.$$

Hence, if $H = \frac{1}{2}$, we have a H–s.a.s.i. Gaussian process, with zero mean and autocovariance function equal to $\min\{s,t\}$. So necessarily it is a standard Brownian motion. However, despite of SMB case, we shall see that a fractional Brownian motion has not generally independent increments (in fact it is the case if and only if $H = \frac{1}{2}$).

A different representation of FBM is due to Mandelbrot and Van Ness (1968). Working on Kolmogorov's elaborates on SBM [22] and Yaglom's ones about processes with stationary increments [42], they obtained a stochastic representation of the (standard) fractional Brownian motion based on Wiener integral. This representation is not unique. As a matter of fact, there exist several representations of fractional Brownian motions as an integral with respect to a Gaussian measure or a Brownian motion. None of them seems to be universal, and the best representation depends on the application at hand. Each of them leads to the construction of new processes. Below we report the Mandelbrot and Van Ness representation, based on a moving average of $\mathrm{d}B(u)$, being the kernel $(t-u)^{H-\frac{1}{2}}$ used as weights.

Given $s < t \in \mathbb{R}$, we have:

$$B_H(t) - B_H(s) = \frac{1}{\Gamma\left(H+\frac{1}{2}\right)} \cdot \left[\int_{-\infty}^{t} (t-u)^{H-\frac{1}{2}}\, \mathrm{d}B(u) - \int_{-\infty}^{s} (s-u)^{H-\frac{1}{2}}\, \mathrm{d}B(u) \right]$$

where $\Gamma(x) = \int_0^{+\infty} t^{x-1} \cdot e^{-x}\, \mathrm{d}x$ is the Euler's Gamma function and $H \in (0,1)$ is the well-known Hurst exponent. The integrator $B(u)$ is a stochastic process, ordinary the Brownian motion. Note that a more general representation can be achieved considering

$$\begin{aligned} \int_{-\infty}^{t} (t-u)^{H-\frac{1}{2}}\, \mathrm{d}B(u) &= \int_{-\infty}^{t} (t-u)^{H-\frac{1}{2}}\, \mathrm{d}B(u) + \int_{t}^{+\infty} 0^{H-\frac{1}{2}}\, \mathrm{d}B(u) = \\ &= \int_{-\infty}^{+\infty} (\max\{t-u,0\})^{H-\frac{1}{2}}\, \mathrm{d}B(u) = \\ &= \int_{\mathbb{R}} (t-u)_+^{H-\frac{1}{2}}\, \mathrm{d}B(u)\,. \end{aligned}$$

So the previous representation can be rewritten as:

$$B_H(t) - B_H(s) = \frac{1}{\Gamma\left(H+\frac{1}{2}\right)} \cdot \int_{\mathbb{R}} \left[(t-u)_+^{H-\frac{1}{2}} - (s-u)_+^{H-\frac{1}{2}} \right] \mathrm{d}B(u)\,.$$

In addiction, since almost surely $B_H(0) = 0$, we have:

$$B_H(t) = \frac{1}{\Gamma\left(H+\frac{1}{2}\right)} \cdot \left[\int_{-\infty}^{t} (t-u)^{H-\frac{1}{2}}\, \mathrm{d}B(u) - \int_{-\infty}^{0} (-u)^{H-\frac{1}{2}}\, \mathrm{d}B(u) \right]. \quad (1.20)$$

Due to this last expression, we can see that, if $H = \frac{1}{2}$, the integral representation leads to the SBM.

$$\begin{aligned} B_{\frac{1}{2}}(t) &= \frac{1}{\Gamma\left(\frac{1}{2}+\frac{1}{2}\right)} \cdot \left[\int_{-\infty}^{t} (t-u)^{\frac{1}{2}-\frac{1}{2}}\, \mathrm{d}B(u) - \int_{-\infty}^{0} (-u)^{\frac{1}{2}-\frac{1}{2}}\, \mathrm{d}B(u)\right] = \\ &= \frac{1}{\Gamma(1)} \cdot \left[\int_{-\infty}^{t} \mathrm{d}B(u) - \int_{-\infty}^{0} \mathrm{d}B(u)\right] = \\ &= \frac{1}{0!} \cdot \left[B(t) - \lim_{u\to-\infty} B(u)\right] - \left[B(0) - \lim_{u\to-\infty} B(u)\right] = \\ &= B(t) - B(0) = B(t)\,. \end{aligned}$$

If we should represent a non-standard FBM (that is $\sigma^2 \neq 1$), the Samorodnitsky and Taqqu's integral representation [38] would suite our propose. It is given by:

$$B_H(t) = \frac{1}{c(H)} \cdot \int_{\mathbb{R}} \left[(t-u)_+^{H-\frac{1}{2}} - (-u)_+^{H-\frac{1}{2}}\right]\, \mathrm{d}B(u) \tag{1.21}$$

where $c(H) = \sqrt{\int_0^{+\infty} \left[(1+u)^{H-\frac{1}{2}} - u^{H-\frac{1}{2}}\right]^2\, \mathrm{d}u + \frac{1}{2H}}$. In this case, it can be proved that $\sigma^2 = c(H)^2 \neq 1$ and that the variance of the process is:

$$\mathbb{V}\left[B_H(t)\right] = c(H)^2 \cdot t^{2H}\,.$$

Let we consider a peculiar feature of FBM, that is, unlike SBM, the absence of independent increments. This implies it may have *infinite memory*. To understand this circumstance, we need to introduce the concept of long-range dependence.

Definition 1.3.5 *A stationary discrete-time processes $\{X_n\}_{n\in\mathbb{N}}$ is said to be a process with* long-range dependence*, long memory, or strong dependence when its autocovariance function*[15] *decays so slowly that*

$$\sum_{\tau=0}^{\infty} \mathcal{C}_X(\tau) = \infty \tag{1.22}$$

in contrast to processes with summable covariances, which are called processes with short-range dependence, short memory, or weak dependence.

Intuitively, when long-range dependence is in place, high-lag correlations may be individually small, but their cumulative effect is significant. Such a covariance structure has an important impact on usual statistical inference.

As an example, assume we have n observations of some random variable X with finite variance. The standard deviation of the mean is then

[15] Since the process is stationary, the mean function is constant and the autocovariance function depends only one variable.

proportional to $\frac{1}{\sqrt{n}}$ if the observations are uncorrelated (so covariances are null).

If the covariances decay exponentially, the covariances are summable and similar behavior is observed: The standard deviation of the mean is proportional to $\frac{1}{\sqrt{n}}$ for sufficiently large n, although the proportionality constant is different.

Suppose, on the other hand, that the covariance function decays hyperbolically[16]

$$\mathcal{C}_X(\tau) \sim \eta \cdot |\tau|^{-\alpha}$$

as $|\tau|$ tends to infinity for some $0 < \alpha < 1$ and a finite positive constant η. It is readily checked that the process is long-range dependent if this relation holds, since the generalized harmonic series diverges for $\alpha < 1$:

$$\sum_{\tau=1}^{\infty} \frac{1}{\tau^{\alpha}} = \infty \,.$$

Under long-range dependence, the standard deviation of the mean is proportional to $\frac{1}{\sqrt{n^{\alpha}}}$. Of course, this affects confidence intervals for the mean of X and all related test statistics. Moreover, the standard estimator for the variance of X becomes biased. This bias does not disappear when the sample size increases, as is the case for short-range dependent processes[17].

Now we can apply the concept of long-range dependence to the fractional Brownian motion.

Theorem 1.3.6 *Given a fractional Brownian motion $\{B_H(t)\}$, if the Hurst exponent is $H \in \left(\frac{1}{2}, 1\right)$, its incremental process exhibits long-range dependence.*

Proof: Define the incremental process $\{X_n\}_{n\in\mathbb{N}}$ of FBM, by

$$X_n := B_H(n+h) - B_H(n) \,.$$

It is clear that $\{X_n\}$ has a standard normal distribution for every n, but that there is (in general) no independence. To be precise, using the property of footnote 2 and with few substitutions, the autocovariance function is of

[16] If $f(x) \sim g(x)$ as $x \to x_0 \in \mathbb{R}^*$, the two functions are asymptotically equivalent. So we have, given $g(x) \neq 0$,

$$\lim_{x \to x_0} \frac{f(x)}{g(x)} = 1 \,.$$

[17] It is important to note that the relation above determines only the *decay* of the correlations. It is still possible that there are some specific lags for which $\mathcal{C}_X(\tau)$ is particularly large, which makes the detecting of long-range dependence more difficult. In fact, it is theoretically impossible to conclude that long-range dependence is present in a finite sample.

the form, where $m+h \leq n$ and $n-m = \tau \cdot h$:

$$\begin{aligned}
\mathcal{C}_X(\tau) &= \mathbb{E}\left(X_n \cdot X_m\right) = \mathbb{E}\big\{[B_H(n+h) - B_H(n)] \cdot [B_H(m+h) - B_H(m)]\big\} = \\
&= \mathcal{C}_{B_H}(n+h, m+h) - \mathcal{C}_{B_H}(n+h, m) - \mathcal{C}_{B_H}(n, m+h) + \mathcal{C}_{B_H}(n, m) = \\
&= \frac{(n+h)^{2H} + (m+h)^{2H} - (n-m)^{2H}}{2} - \frac{(n+h)^{2H} + m^{2H} - (n+h-m)^{2H}}{2} - \\
&\quad - \frac{n^{2H} + (m+h)^{2H} - (n-h-m)^{2H}}{2} + \frac{n^{2H} + m^{2H} - (n-m)^{2H}}{2} = \\
&= \frac{(n-m+h)^{2H} + (n-m-h)^{2H} - 2 \cdot (n-m)^{2H}}{2} = \\
&= \frac{(\tau \cdot h + h)^{2H} + (\tau \cdot h - h)^{2H} - 2 \cdot (\tau \cdot h)^{2H}}{2} = \\
&= \frac{[h \cdot (\tau+1)]^{2H} + [h \cdot (\tau-1)]^{2H} - 2 \cdot (\tau \cdot h)^{2H}}{2} = \\
&= h^{2H} \cdot \frac{(\tau+1)^{2H} + (\tau-1)^{2H} - 2\tau^{2H}}{2} .
\end{aligned}$$

Clearly, if $H = \frac{1}{2}$, the covariances are null (except, of course, for $h = 0$), collapsing the process towards a SMB. Since the process is still Gaussian, uncorrelation implies independence.

Consider now the function $f(\tau) = (1+\tau)^{2H} + (1-\tau)^{2H} - 2$ and compute its Maclaurin series to second order. Since its derivatives are given by $f^{(k)}(\tau) = \prod_{i=0}^{k-1}(2H-i) \cdot \left[(1+\tau)^{2H-k} + (-1)^k \cdot (1-\tau)^{2H-k}\right]$, the approximation is:

$$f(\tau) \approx f(0) + f'(0) \cdot \tau + \frac{f''(0)}{2} \cdot \tau^2 \approx 2H \cdot (2H-1) \cdot \tau^2 .$$

And since:

$$\begin{aligned}
\tau^{2H} \cdot f\left(\frac{1}{\tau}\right) &= \tau^{2H} \cdot \left[\left(1 + \frac{1}{\tau}\right)^{2H} + \left(1 - \frac{1}{\tau}\right)^{2H} - 2\right] = \\
&= \left[\tau \cdot \left(1 + \frac{1}{\tau}\right)\right]^{2H} + \left[\tau \cdot \left(1 - \frac{1}{\tau}\right)\right]^{2H} - 2\tau^{2H} = \\
&= (\tau+1)^{2H} + (\tau-1)^{2H} - 2\tau^{2H} .
\end{aligned}$$

Due to the last two results, the autocovariance function can be approximated by:

$$h^{2H} \cdot \frac{\tau^{2H} \cdot f\left(\frac{1}{\tau}\right)}{2} \approx h^{2H} \cdot \frac{\tau^{2H} \cdot 2H \cdot (2H-1) \cdot \left(\frac{1}{\tau}\right)^2}{2} \approx h^{2H} \cdot \left[H \cdot (2H-1) \cdot \tau^{2H-2}\right] .$$

Hence, since $0 < H < 1$, as τ tends to infinity:

$$\mathcal{C}_X(\tau) \sim h^{2H} \cdot H \cdot (2H-1) \cdot \tau^{2H-2} .$$

Therefore:

- If $\frac{1}{2} < H < 1$ (divergent generalized harmonic series), $\mathcal{C}_X(\tau)$ is positive with

$$\sum_{\tau=0}^{\infty} \mathcal{C}_X(\tau) = \infty\,;$$

- If $0 < H < \frac{1}{2}$ (convergent generalized harmonic series or Riemann zeta function), $\mathcal{C}_X(\tau)$ is negative with

$$\sum_{\tau=0}^{\infty} |\,\mathcal{C}_X(\tau)| < \infty\,.$$

□

On one hand, in the first case the sequence shows an aggregative behavior which can be used to describe cluster phenomena. On the other hand, if there is no long-range dependence, such a process can be used to model sequences with intermittency. In the following presentation, we will be more interested in the case $\frac{1}{2} < H < 1$.

The last issue of this section is to show that FBM is not a semimartingale and hence a martingale neither.

Theorem 1.3.7 *Given a fractional Brownian motion* $\{B_H(t)\}_{t\,\in\,[0,+\infty)}$*, unless* $H = \frac{1}{2}$*, it is not a semimartingale.*

Proof: Fix $p > 0$. We define the p–variation of a continuous stochastic process $\{Y(t)\}$, as the following limit in probability[18]

$$\operatorname*{plim}_{n\to\infty} \sum_{j=1}^{n} \left| Y\left(\frac{j}{n}\right) - Y\left(\frac{j-1}{n}\right)\right|^p .$$

If the p–variation exists and it is nonzero, then for every $q > p$ the p–variation is zero and for any $q < p$ the p–variation is infinite.
Let we consider the a FBM. By the self-affinity property[19], the discrete

[18] A sequence $\{X_n\}_{n\,\in\,\mathbb{N}}$ of random variables converges in probability towards the variable X if for all $\varepsilon > 0$:

$$\lim_{n\to\infty} \Pr\{|X_n - X| \geq \varepsilon\} = 0\,.$$

Convergence in probability is denoted by adding the letter P over an arrow indicating convergence, or using the probability limit operator:

$$X_n \xrightarrow{P} X \quad \text{or} \quad \operatorname*{plim}_{n\to\infty} X_n = X\,.$$

[19] If $\{X(t)\}$ is a H–s.a. process, for the process $\{X(t)^p\}$ it ensues:

$$\{X(a\cdot t)^p\} \overset{d}{\sim} \left\{\left[a^H \cdot X(t)\right]^p\right\} = \left\{a^{pH} \cdot X(t)^p\right\}\,.$$

stochastic process

$$X_n := n^{pH-1} \cdot \sum_{j=1}^{n} \left| B_H\left(\frac{j}{n}\right) - B_H\left(\frac{j-1}{n}\right) \right|^p$$

has the same distribution as

$$\tilde{X}_n := n^{pH-1} \cdot \left(\frac{1}{n}\right)^{pH} \cdot \sum_{j=1}^{n} |B_H(j) - B_H(j-1)|^p =$$

$$= \frac{1}{n} \cdot \sum_{j=1}^{n} |B_H(j) - B_H(j-1)|^p .$$

By the Ergodic Theorem $\tilde{X}_n$ converges in L^1 to $\mathbb{E}\left(|B_H(1)|^p\right) \neq 0$ as n tends to infinity. As a consequence, since convergence in L^1 implies convergence in probability, X_n converges in probability as n tends to infinity to the same value. Therefore:

$$\operatorname*{plim}_{n\to\infty} \sum_{j=1}^{n} \left| B_H\left(\frac{j}{n}\right) - B_H\left(\frac{j-1}{n}\right) \right|^p = \begin{cases} 0 & \text{if } pH > 1 \\ \infty & \text{if } pH < 1 . \end{cases}$$

As a consequence, the FBM with Hurst exponent $H \neq \frac{1}{2}$ is not a semimartingale. In fact, semimartingales can be expressed as the sum of a bounded variation process and a local martingale which has finite quadratic variation. The FBM cannot be a semimartingale except in the case $H = \frac{1}{2}$ because:

1. If $H < \frac{1}{2}$, we can choose $p > 2$ such that $pH < 1$ and we obtain that the p–variation of FBM is infinite. Hence, the quadratic variation ($p = 2$) is also infinite.

2. If $H > \frac{1}{2}$, we can choose p such that $\frac{1}{H} < p < 2$. Then the p–variation is zero, and, as a consequence, the quadratic variation is also zero. On the other hand, if we choose p such that $1 < p < \frac{1}{H}$ we deduce that the total variation is infinite.

Therefore, we have proved that for $H \neq \frac{1}{2}$ FBM cannot be a semimartingale.

□

1.4 Pareto's distributions

In this section we introduce a very peculiar distribution which will be used at a later stage to characterize a more general class of random variables. This distribution owes its name to the scientist who dealt with it firstly. He was the Italian economist Vilfredo Pareto, well-know for his law about income

distribution[20]. If a random variable follows the Pareto's distribution, it means that its right tail behaves like a *power-law*; for this reason the Pareto's distribution is even called *power-law distribution*.

We define a power-law such a function $f : \mathbb{R} \to \mathbb{R}$ such that:

$$f(x) := C \cdot x^k + o(x^k) \tag{1.23}$$

where C and k are constant, the latter named *scaling exponent*. As a matter of fact, the most important property of these functions is to be asymptotically invariant under all rescalings[21], since it follows:

$$\lim_{x \to \pm\infty} f(\lambda \cdot x) = \lim_{x \to \pm\infty} C \cdot (\lambda \cdot x)^k = \lambda^k \cdot \lim_{x \to \pm\infty} C \cdot x^k = \lambda^k \cdot \lim_{x \to \pm\infty} f(x)\,.$$

We shall see that scale invariance is deeply connected with self-affinity.

The main issue of this section is to show how Pareto's distribution have to be built, and hence to underline few of its properties. To have a power-law distribution we need a continuous random variable whose support is the positive real line and *tail distribution* such that, given $C \in \mathbb{R}^+$:

$$\Pr\{X > x\} = C \cdot x^{-\alpha}\,. \tag{1.24}$$

It means that its cumulative distribution function is given by:

$$F_X(x) = \Pr\{X \leq x\} = 1 - \Pr\{X > x\} = 1 - C \cdot x^{-\alpha}\,.$$

So defined, the random variable is ill posed. As a matter of fact, every c.d.f. must follow the two conditions:

$$\lim_{x \to 0^+} F_X(x) = 0 \qquad \text{and} \qquad \lim_{x \to +\infty} F_X(x) = 1\,.$$

Without setting further restrictions over the two parameters C and α, the two limits might diverge. Indeed:

$$\lim_{x \to 0^+} F_X(x) = \lim_{x \to 0^+} 1 - \frac{C}{x^\alpha} = \begin{cases} -\infty & \text{if } \alpha > 0 \\ 1 - C & \text{if } \alpha = 0 \\ 1 & \text{if } \alpha < 0 \end{cases}$$

and

[20] He noticed that that 80% of Italy's land was owned by 20% of the population. He then carried out surveys on a variety of other countries and found to his surprise that a similar distribution applied. So he hypothesized the *80–20 rule*: Twenty percent of global population owns the eighty percent of the global wealth.

[21] A function is said to be invariant under scaling if and if only:

$$f(\lambda \cdot x) = \lambda^\beta \cdot f(x)$$

for some choice of exponent β, and and for all dilations $\lambda > 0$. This is equivalent to f being a homogeneous function.

$$\lim_{x\to+\infty} F_X(x) = \lim_{x\to+\infty} 1 - \frac{C}{x^\alpha} = \begin{cases} 1 & \text{if } \alpha > 0 \\ 1 - C & \text{if } \alpha = 0 \\ -\infty & \text{if } \alpha < 0\,. \end{cases}$$

We see how the first limit never becomes null. For the second one we have to set $\alpha > 0$. However the "problem" related to the first limit cannot be solved, unless modifying it from the beginning. A obvious reply is to threshold the support of the random variable to a subset of the positive real line. Hence we can set $x \in [x_{min}, +\infty) \subset \mathbb{R}^+$, in order to find the value of C such that:

$$\lim_{x\to x_{min}^+} 1 - \frac{C}{x^\alpha} = 0\,.$$

Since the c.d.f. is continuous in $x = x_{min}$, we have:

$$C = \left(\frac{1}{x_{min}}\right)^{-\alpha}\,. \tag{1.25}$$

The fact of restricting the support in a neighborhood of the origin does not entail great changes or loss of generality. As a matter of fact, Pareto's distribution is almost always chosen for the property of its (heavy) *tails*. What happens at the beginning of the domain is not so crucial.

We are now able to define the Pareto's distribution properly.

Definition 1.4.1 *Given a real valued random variable defined on the support $[x_{min}, +\infty) \subset \mathbb{R}^+$, we say it is Paretian distributed if and if only its cumulative distribution function is so defined*

$$F_X(x) := 1 - \left(\frac{x}{x_{min}}\right)^{-\alpha} \tag{1.26}$$

where α is a real positive (scaling) parameter. To say that a random variable follows the Pareto's distribution we write $X \overset{d}{\sim} \mathcal{P}(\alpha, x_{min})$.

Below we deduce its probability density function

$$f_X(x) = F'_X(x) = \frac{\alpha}{x} \cdot \left(\frac{x}{x_{min}}\right)^{-\alpha}$$

and its moments. In particular it results very important to understand for

what values of α they exist and are finite. The first moment is:

$$
\begin{aligned}
m_1 = \mathbb{E}(X) &= \int_{x_{min}}^{+\infty} x \cdot f_X(x)\,\mathrm{d}x = \\
&= \int_{x_{min}}^{+\infty} x \cdot \frac{\alpha}{x} \cdot \left(\frac{x}{x_{min}}\right)^{-\alpha}\,\mathrm{d}x = \\
&= \alpha \cdot x_{min}^{\alpha} \int_{x_{min}}^{+\infty} x^{-\alpha}\,\mathrm{d}x = \\
&= \begin{cases} \alpha \cdot x_{min}^{\alpha} \cdot \left(\frac{x^{1-\alpha}}{1-\alpha}\right)\Big|_{x_{min}}^{+\infty} & \text{if } \alpha \neq 1 \\ \\ \alpha \cdot x_{min}^{\alpha} \cdot (\ln|x|)\Big|_{x_{min}}^{+\infty} & \text{if } \alpha = 1 \end{cases} \\
&= \begin{cases} \alpha \cdot x_{min}^{\alpha} \cdot \left(\lim_{x\to+\infty} \frac{x^{1-\alpha}}{1-\alpha} - \frac{x_{min}^{1-\alpha}}{1-\alpha}\right) & \text{if } \alpha \neq 1 \\ \\ \alpha \cdot x_{min}^{\alpha} \cdot \left(\lim_{x\to+\infty} \ln|x| - \ln|x_{min}|\right) & \text{if } \alpha = 1 \end{cases} \\
&= \begin{cases} \infty & \text{if } 0 < \alpha \leq 1 \\ \frac{\alpha}{\alpha-1} \cdot x_{min} & \text{if } \alpha > 1\,. \end{cases}
\end{aligned}
$$

The second moment is given by:

$$
\begin{aligned}
m_2 = \mathbb{E}\big(X^2\big) &= \int_{x_{min}}^{+\infty} x^2 \cdot f_X(x)\,\mathrm{d}x = \\
&= \int_{x_{min}}^{+\infty} x^2 \cdot \frac{\alpha}{x} \cdot \left(\frac{x}{x_{min}}\right)^{-\alpha}\,\mathrm{d}x = \\
&= \alpha \cdot x_{min}^{\alpha} \int_{x_{min}}^{+\infty} x^{1-\alpha}\,\mathrm{d}x = \\
&= \begin{cases} \alpha \cdot x_{min}^{\alpha} \cdot \left(\frac{x^{2-\alpha}}{2-\alpha}\right)\Big|_{x_{min}}^{+\infty} & \text{if } \alpha \neq 2 \\ \\ \alpha \cdot x_{min}^{\alpha} \cdot (\ln|x|)\Big|_{x_{min}}^{+\infty} & \text{if } \alpha = 2 \end{cases} \\
&= \begin{cases} \alpha \cdot x_{min}^{\alpha} \cdot \left(\lim_{x\to+\infty} \frac{x^{2-\alpha}}{2-\alpha} - \frac{x_{min}^{2-\alpha}}{2-\alpha}\right) & \text{if } \alpha \neq 2 \\ \\ \alpha \cdot x_{min}^{\alpha} \cdot \left(\lim_{x\to+\infty} \ln|x| - \ln|x_{min}|\right) & \text{if } \alpha = 2 \end{cases} \\
&= \begin{cases} \infty & \text{if } 0 < \alpha \leq 2 \\ \frac{\alpha}{\alpha-2} \cdot x_{min}^2 & \text{if } \alpha > 2\,. \end{cases}
\end{aligned}
$$

All these computations proves that the existence of the moments depends on

the value assumed by α[22]. In particular we are interested in understanding under which values the variance becomes infinite. That is:

$$\mathbb{V}(X) = m_2 - m_1^2 = \begin{cases} \infty & \text{if } 0 < \alpha \leq 2 \\ \frac{\alpha}{(\alpha-1)^2 \cdot (\alpha-2)} \cdot x_{min}^2 & \text{if } \alpha > 2\,. \end{cases} \tag{1.27}$$

In the following we will focus mostly on Pareto's distribution having the scaling parameter $\alpha \in (1, 2]$, since they have finite mean but infinite variance. A first implication of such distributions is the fact that the Central Limit Theorem is not a useful approximation, since the theoretical assumptions for it do not hold. Drawing a random sample from a Paretian distribution repeatedly and average its elements, the mean we get is neither normally distributed nor satisfies a Law of Large Numbers. Things do not "average out". If applied to financial data, this circumstance can be traced when, computing the sample variance for different time horizons, it seems to diverge. Subsequently we will focus on it in depth.

Next we show that the Pareto's distribution is invariant under scaling. It immediate to see that:

$$f_X(\lambda \cdot x) = \frac{\alpha}{\lambda \cdot x} \cdot \left(\frac{\lambda \cdot x}{x_{min}}\right)^{-\alpha} = \lambda^{-\alpha-1} \cdot \frac{\alpha}{x} \cdot \left(\frac{x}{x_{min}}\right)^{-\alpha} = \lambda^{-\alpha-1} \cdot f_X(x)\,.$$

Thus the p.d.f. is a homogeneous function of degree $-\alpha - 1$. In addiction such a random variable is often plotted in a linear scale. In order to do that,

[22] More generally, the k-th moment is:

$$\begin{aligned}
m_k = \mathbb{E}(X^k) &= \int_{x_{min}}^{+\infty} x^k \cdot f_X(x)\, \mathrm{d}x = \\
&= \int_{x_{min}}^{+\infty} x^k \cdot \frac{\alpha}{x} \cdot \left(\frac{x}{x_{min}}\right)^{-\alpha} \mathrm{d}x = \\
&= \alpha \cdot x_{min}^{\alpha} \int_{x_{min}}^{+\infty} x^{k-1-\alpha}\, \mathrm{d}x = \\
&= \begin{cases} \alpha \cdot x_{min}^{\alpha} \cdot \left(\frac{x^{k-\alpha}}{k-\alpha}\right) \Big|_{x_{min}}^{+\infty} & \text{if } \alpha \neq k \\ \\ \alpha \cdot x_{min}^{\alpha} \cdot (\ln|x|) \Big|_{x_{min}}^{+\infty} & \text{if } \alpha = k \end{cases} \\
&= \begin{cases} \alpha \cdot x_{min}^{\alpha} \cdot \left(\lim_{x \to +\infty} \frac{x^{k-\alpha}}{k-\alpha} - \frac{x_{min}^{k-\alpha}}{k-\alpha}\right) & \text{if } \alpha \neq k \\ \\ \alpha \cdot x_{min}^{\alpha} \cdot \left(\lim_{x \to +\infty} \ln|x| - \ln|x_{min}|\right) & \text{if } \alpha = k \end{cases} \\
&= \begin{cases} \infty & \text{if } 0 < \alpha \leq k \\ \frac{\alpha}{\alpha-k} \cdot x_{min}^k & \text{if } \alpha > k\,. \end{cases}
\end{aligned}$$

it is sufficient to take the logarithms of its p.d.f. for $x \geq x_{min}$, that is

$$\ln\left[f_X(x)\right] = \ln\left[\frac{\alpha}{x}\cdot\left(\frac{x}{x_{min}}\right)^{-\alpha}\right] = \ln\left(\alpha\cdot x_{min}^{\alpha}\cdot x^{-\alpha-1}\right) =$$
$$= \ln\left(\alpha\cdot x_{min}^{\alpha}\right) + \ln\left(x^{-\alpha-1}\right) = \ln\left(\alpha\cdot x_{min}^{\alpha}\right) - (\alpha+1)\cdot\ln(x)\ .$$

Since α is positive, we obtain a straight line with negative slope and positive intercept $\ln\left(\alpha\cdot x_{min}^{\alpha}\right)$. This graphical representation is called *log-log plot*, since we take both the logarithm of the p.d.f. and the logarithm of the random variable's values. So, if we believe a random variable has got *heavy tails*, its log-log plot should be approximatively as a straight line with negative gradient (see also Definition 1.5.7).

Furthermore, this method is even a very useful way to distinguish between a given distribution (for instance a Paretian) and some other distribution. Take for instance a exponential distribution $X \overset{d}{\sim} \mathcal{E}(\lambda)$:

$$\ln\left[f_X(x)\right] = \ln\left(\lambda\cdot e^{-\lambda\cdot x}\right) = \ln(\lambda) + \ln\left(e^{-\lambda\cdot x}\right)$$
$$= \ln(\lambda) - \lambda\cdot x = \ln(\lambda) - \lambda\cdot e^{\ln(x)}\ .$$

Thus its graph, on the plane $(\ln(f), \ln(x))$, is a concave curve. The same is for the the log-normal distribution $X \overset{d}{\sim} \mathcal{LN}(\mu, \sigma^2)$. Its log-log graph is:

$$\ln\left[f_X(x)\right] = \ln\left\{\frac{1}{x\sqrt{2\cdot\pi\cdot\sigma^2}}\cdot e^{-\frac{[\ln(x)-\mu]^2}{2\cdot\sigma^2}}\right\} =$$
$$= \ln\left[\frac{1}{x\sqrt{2\cdot\pi\cdot\sigma^2}}\cdot e^{-\frac{\ln(x)^2-2\cdot\mu\cdot\ln(x)+\mu^2}{2\cdot\sigma^2}}\right] =$$
$$= -\ln(x) - \frac{\ln\left(2\cdot\pi\cdot\sigma^2\right)}{2} - \frac{\ln(x)^2 - 2\cdot\mu\cdot\ln(x) + \mu^2}{2\cdot\sigma^2} =$$
$$= -\frac{1}{2\cdot\sigma^2}\cdot\ln(x)^2 + \left(\frac{\mu}{\sigma^2} - 1\right)\cdot\ln(x) - \frac{\mu^2 + \sigma^2\cdot\ln\left(2\cdot\pi\cdot\sigma^2\right)}{2\cdot\sigma^2}\ .$$

This one is a quadratic function of $\ln(x)$ with negative curvature, so clearly distinguishable from a straight line. However, all these three distribution are linked in a particular way. As a matter of fact, consider a $X \overset{d}{\sim} \mathcal{P}(\alpha, x_{min})$. Setting

$$Y := \ln\left(\frac{X}{x_{min}}\right)$$

its cumulative distribution function is:

$$F_Y(y) = \Pr\{Y \leq y\} = \Pr\left\{\ln\left(\frac{X}{x_{min}}\right) \leq y\right\} = \Pr\left\{\frac{X}{x_{min}} \leq e^y\right\} =$$
$$= \Pr\left\{X \leq x_{min}\cdot e^y\right\} = 1 - \left(\frac{x_{min}\cdot e^y}{x_{min}}\right)^{-\alpha} = 1 - e^{-\alpha\cdot y}\ .$$

Since the last expression is the c.d.f. of a random variable exponentially distribute, hence $Y \overset{d}{\sim} \mathcal{E}(\alpha)$. This show how Paretian distribution and log-normal distribution are alternative distributions for describing the same types of quantities. One of the connections between the two is that they are both the distributions of the exponential of random variables distributed according to other common distributions, respectively the exponential distribution and normal distribution[23].

1.5 Lévy-stable random variables

We introduce a fundamental class of random variables which allows to generalize the Central Limit Theorem in the case of infinite variance. That class of variables usually takes the name of the French mathematician Paul Lévy who gave key contributions to the emergence of the probabilistic version of the notion of *scaling*, as implemented in the probability distributions that he called *stable*. Let we start with a classic result that we will not prove.

Theorem 1.5.1 (Lindeberg–Lévy Central Limit Theorem) *Consider a sequence of identically and independently distributed random variables* $X_1, X_2, \ldots, X_n$. *If the first two moments exist and are finite, the corresponding sequence of the normalized sums*

$$Z_n = \frac{\sum_{i=1}^{n} X_i - n \cdot \mathbb{E}(X)}{\sqrt{n \cdot \mathbb{V}(X)}} \tag{1.28}$$

converge in distribution[24] *to a standard normal random variable.*

The Central Limit Theorem has been extended in various directions. One of them was aimed to extend the understanding of the Central Limit Theorem with the use of not only the normal law as a limiting approximation but also some other distributions of a certain analytical structure.

The formulation is as follows. A sequence of independent and identically distributed random variables $X_1, X_2, \ldots, X_n$ is considered, without any preliminary assumptions about their distribution. With the use of sequences of real-valued constants $a_1, a_2, \ldots, a_n$ and positive constants $b_1, b_2, \ldots, b_n$ the standardized sums

$$Z_n = \frac{\sum_{i=1}^{n} X_i - a_n}{b_n}$$

[23] As a matter of fact

$$Y \overset{d}{\sim} \mathcal{N}(\mu, \sigma^2) \Leftrightarrow e^Y \overset{d}{\sim} \mathcal{LN}(\mu, \sigma^2) \qquad \text{and} \qquad Y \overset{d}{\sim} \mathcal{E}(\alpha) \Leftrightarrow x_{min} \cdot e^Y \overset{d}{\sim} \mathcal{P}(\alpha, x_{min}) .$$

[24] A sequence of random variables $X_1, X_2, ..., X_n$ is said to converge in distribution random variable X if $\lim_{n\to\infty} F_{X_n}(x) = F_X(x)$ for every $x \in \mathbb{R}$ at which F is defined and continuous. Of course, $F_n(x)$ and $F(x)$ are the cumulative distribution functions of random variables X_n and X, respectively. It is commonly written as $X_n \overset{d}{\to} X$.

are introduced. We see that in the classic Central Limit Theorem we fixed $a_n = n \cdot \mathbb{E}(X)$ and $b_n = \sqrt{n \cdot \mathbb{V}(X)}$. We assume now that the constants a_n and b_n are chosen in an appropriate way so that the distribution functions of Z_n converge in distribution to some limit distribution function $F_X(x)$, that is,

$$Z_n \xrightarrow{d} X$$

for any x which is a continuity point of the function $F_X(x)$.

A problem thus arises: How wide is the class of the distribution functions that can play the role of the limiting law? The answer to this question define the class of *stable* (more specifically, Lévy-stable or L-stable) random variables.

Definition 1.5.2 *A random variable X is referred to as* stable *if there exist constants a_n and $b_n > 0$ such that*

$$\sum_{i=1}^{n} X_i \overset{d}{\sim} a_n + b_n \cdot X \tag{1.29}$$

where X_1, X_2, ..., X_n are independent random variables each having the same distribution as X. Moreover a variable is said to be strictly stable *if there exist constants $b_n > 0$ such that*

$$\sum_{i=1}^{n} X_i \overset{d}{\sim} b_n \cdot X\,. \tag{1.30}$$

If in the standardized sums the random variables X_i are supposed to be equal to one and the same constant c with probability one, and the normalizing constants are $a_n = c \cdot (n-1)$ and $b_n = 1$, then the sums Z_n for any n also turn out to be equal to the constant c with probability one. In fact:

$$Z_n = \sum_{i=1}^{n} c - c \cdot (n-1) = n \cdot c - n \cdot c + c = c\,.$$

The distribution function of such random variables is called *degenerate* at the point c and its graph resembles the form of the step function. Moreover, a random variable taking a single value with probability one does not depend on any other random variable. Therefore, all the random variables X_i in the case under consideration are mutually independent and the limit distribution for the sums Z_n, by definition, is to belong to the stable law family. In other words, the family includes all degenerate distribution functions.

However, to give a non trivial answer to the previous question, the problem appears to be not so simple to be solved. The first results concerning this problem were obtained by the Hungarian mathematician György Pólya. Among the facts he collected, one is worthwhile to notice: Only the normal law (excepting, of course, the degenerate law) is the stable random variable

that possesses a finite variance; the variance of any other non-degenerate stable distribution is infinite.

The next considerable step forward was made by the French mathematician Paul Lévy in 1925. In his book *Calcul des Probabilites*, a special chapter was devoted to this problem. He introduced stable laws and the concept of stability, moreover specifying the further subclass of the so-called strictly stable laws. Furthermore he revealed a functional equation which the strictly stable laws should satisfy. More specifically, its distribution function is the convolution of the distribution functions of the summands[25], that is

$$F_S(x) = F_{X_1}\left(\frac{x}{C_1}\right) * F_{X_2}\left(\frac{x}{C_2}\right) \quad \Longrightarrow \quad F_X(x) = F_X\left(\frac{x}{C_1}\right) * F_X\left(\frac{x}{C_2}\right)$$

where F_X is the c.d.f. that the standardized sums should converge towards. Also C_1 and C_2 are two positive constants such that $C_1^\alpha + C_2^\alpha = \gamma^\alpha$ and $\alpha \in (0, 2]$ is the characteristic parameter of the stable law (see Proposition 1.5.3).

The description of the remaining part of the stable laws was finished more than a decade later in the middle of thirties. It was made by Lévy himself and the Soviet mathematician Aleksandr Yakovlevich Khinchin independently of each other and almost simultaneously. The functional equation whose solutions form the family of stable laws was not difficult to derive; it is obtained in the same way as functional equation for strictly stable laws, and little differs from it:

$$F_X(x + \delta) = F_X\left(\frac{x}{C_1}\right) * F_X\left(\frac{x}{C_2}\right)$$

where δ is a real-valued constant.

In consideration of all of the above, we can now present their explicit formulation. All the stable laws, except the degenerate ones, are absolutely continuous; the corresponding distribution functions $F_X(x)$ possess the densities $f_X(x)$ but, for a few exceptions, *neither* the c.d.f *nor* the p.d.f. can be explicitly expressed in terms of elementary functions. The stable laws are adequately described in terms of the corresponding characteristic functions:

$$\varphi_X(u) = \int_{-\infty}^{+\infty} e^{\imath\, u\, x}\, f_X(x)\, \mathrm{d}x\,.$$

Moreover, looking at the functional equations above, they become simpler in terms of characteristic functions. For example, the more general one becomes:

$$\varphi_X(u) \cdot e^{-\imath\, \delta\, u} = \varphi_X(C_1 \cdot u) \cdot \varphi_X(C_2 \cdot u)\,.$$

While solving this equation, it appears that the set of its solution is a four-parametric family of functions. Indeed each stable law is defined by four parameters.

[25] We have $X_1 \overset{d}{\sim} X_2$ and $S = C_1 \cdot X_1 + C_2 \cdot X_2$. Moreover the symbol $*$ indicates the convolution operation.

Proposition 1.5.3 *Given four parameters $\alpha, \beta, \gamma, \delta$ such that*

- $\alpha \in \mathbb{R} : 0 < \alpha \leq 2$ (*index of stability or shape parameter*);
- $\beta \in \mathbb{R} : -1 \leq \beta \leq 1$ (*skewness parameter*);
- $\gamma \in \mathbb{R} : \gamma > 0$ (*scale parameter*);
- $\delta \in \mathbb{R}$ (*location parameter*);

we say that a random variable is L-stable *if it has a characteristic function which can be written in the following form:*

$$\varphi_X(u) = \begin{cases} e^{\imath\,\delta\,u - \gamma^\alpha\,|u|^\alpha \cdot \left[1 + \imath\,\beta \cdot \operatorname{sgn}(u) \cdot \tan\left(\alpha \cdot \frac{\pi}{2}\right)\right]} & \text{if } \alpha \neq 1 \\ e^{\imath\,\delta\,u - \gamma\,|u| \cdot \left[1 + \frac{2\,\imath\,\beta}{\pi} \cdot \operatorname{sgn}(u) \cdot \ln|u|\right]} & \text{if } \alpha = 1\,. \end{cases} \tag{1.31}$$

In a more compact way, we write:

$$X \overset{d}{\sim} \mathcal{S}\left(\alpha, \beta, \gamma, \delta\right)\,.$$

There, α and β determine the shape of the distribution, while γ and δ determine scale and location in the standard way. More specifically, α, the most important parameter, causes the heaviness of the *tails* of the distribution; β is responsible of the degree of *asymmetry*; γ is the *scale* parameter, that is it determines the entity of the overall probabilities; δ is a *positioning* parameter (it is *like* a mean). Moreover, the distribution is said to be *standardized* if $X \overset{d}{\sim} \mathcal{S}\left(\alpha, \beta, 1, 0\right)$. In fact, given $\alpha \neq 1$, if $X \overset{d}{\sim} \mathcal{S}\left(\alpha, \beta, \gamma, \delta\right)$, then

$$\frac{X - \delta}{\gamma} \overset{d}{\sim} \mathcal{S}\left(\alpha, \beta, 1, 0\right)\,.$$

Before examining in depth the whole subject, we need to underline that the Levy-stable class of random variables is the widest one which make possible to find a law of standardized sums. Moreover, it is obvious, that the normal distribution is a stable law (we will show explicitly later). However, the analysis of the present literature shows that there are only a few dozens of problems, mostly in physics and the related sciences, where the stable laws have been applied, unlike normal one which *seems* to be applicable almost in every topics. Few reason for this can be found in the following list (for a more detailed dissertation see [40]):

1. The first reason, absolutely inessential from a mathematical viewpoint but of great importance for those who use this technique to solve his own particular problems, consists of the absence of simple analytical expressions for the densities of stable laws. That is the reason why the normal law is the first of stable laws which appeared on the scene and found at once the applications in the theory of errors. The next stable laws that appeared in theoretical physics was the Cauchy distribution, which has a closed-form probability density function as well.

2. The second reason consists in that all non-degenerate stable laws, differing from the normal one, have infinite variance. As concerns the whole family of stable laws, the normal law is exceptional, something like a *monster*, an abnormal member of the family. However, from the point of view of the applied sciences, where a specific meaning is often associated with the variance, distributions with infinite variance look odd indeed. Mandelbrot in [27] and [28] wrote that to anyone with the usual training in statistics, an infinite variance seems at best scary and at worst bizarre. Also, of course, the fact that a variable X has an infinite variance in no way denies that X is finite with a probability equal to one. Thus, the choice between variables should not be decided *a priori*, and should hinge solely on the question of which is *more convenient* to handle. In some cases, we should accept infinite variance because it is the only way to preserve *scaling*.

3. Finally, the possibly most trivial but, at the same time, the most important reason for the slow expansion of stable laws in applied problems is the shortage of knowledge. Up to nowadays, as a rule, the information about the stable laws is absent in the textbooks on the Probability Theory. Excluding the Lévy work, we have found only two monographs ([20] and [40]), or single chapters of textbook of advanced Probability (such as in [18]).

In the following part we will prove that some well-known distribution (normal, Cauchy's and Levy's) are particular cases of the more comprehensive class of stable ones.

(a) $\alpha = 2$ and $\beta \in [-1, 1]$ (*Gaussian distribution*)

The characteristic function becomes:

$$\varphi_X(u) = e^{\imath\,\delta\,u - \gamma^2\,|u|^2 \cdot \left[1 + \imath\,\beta \cdot \operatorname{sgn}(u) \cdot \tan\left(2 \cdot \frac{\pi}{2}\right)\right]} = e^{\imath\,\delta\,u - \gamma^2\,u^2}\,.$$

Setting $\gamma^2 = \frac{\sigma^2}{2}$ and $\delta = \mu$, we have:

$$\varphi_X(u) = e^{\imath\,\mu\,u - \frac{\sigma^2\,u^2}{2}}$$

which is the characteristic function of a normal random variable $X \overset{d}{\sim} \mathcal{N}\left(\mu, \sigma^2\right)$.

(b) $\alpha = 1$ and $\beta = 0$ (*Cauchy's distribution*)

The characteristic function becomes:

$$\varphi_X(u) = e^{\imath\,\delta\,u - \gamma\,|u| \cdot \left[1 + \frac{2\,\imath\,0}{\pi} \cdot \operatorname{sgn}(u) \cdot \ln|u|\right]} = e^{\imath\,\delta\,u - \gamma\,|u|}\,.$$

Setting $\gamma = \theta_2$ and $\delta = \theta_1$, we have:

$$\varphi_X(u) = e^{\imath\,\theta_1\,u - \theta_2\,|u|}$$

which is the characteristic function of a random variable following the Cauchy's distribution $X \overset{d}{\sim} \mathcal{C}a\,(\theta_1, \theta_2)$.

(c) $\alpha = \frac{1}{2}$ and $\beta = \pm 1$ (*Lévy's distribution*)

The characteristic function becomes:

$$\varphi_X(u) = e^{\imath\,\delta\,u - \gamma^{\frac{1}{2}}\,|u|^{\frac{1}{2}}\cdot\left[1+\imath\cdot(\pm 1)\cdot\operatorname{sgn}(u)\cdot\tan\left(\frac{1}{2}\cdot\frac{\pi}{2}\right)\right]} = e^{\imath\,\delta\,u - \gamma^{\frac{1}{2}}\,|u|^{\frac{1}{2}}\cdot\left[1\pm\imath\cdot\operatorname{sgn}(u)\right]}\,.$$

Setting $\gamma = c$ and $\delta = \mu$, we have:

$$\varphi_X(u) = e^{\imath\,\mu\,u - \sqrt{c\,|u|}\cdot\left[1\pm\imath\cdot\operatorname{sgn}(u)\right]}$$

which is the characteristic function of a Lévy's random variable $X \overset{d}{\sim} \mathcal{L}e\,(\mu, c)$, where $x \in \mathbb{R} : x > \mu$.

With the exception of the Gauss, Cauchy and Lévy laws, explicit expressions for the densities of stable distributions in terms of elementary functions are unknown, and for their use it is necessary to refer to tables (such as [11]) or use numerical algorithms.

In order to show a more general Central Limit Theorem, we firstly analyse strictly stable random variables determining the positive coefficients b_n that allows for stability.

Theorem 1.5.4 *: For a strictly stable random variable X, the condition*

$$b_n = n^{\geq \frac{1}{\alpha}} \tag{1.32}$$

holds, where α is the index of stability.

Proof: We have to find $b_n > 0$ such that

$$\sum_{i=1}^{n} X_i \overset{d}{\sim} b_n \cdot X$$

This problem is most easily solved for the normal distribution (which is the only stable distribution with finite variance). Since X_1, X_2, ..., X_n are independent random variables each having the same distribution as X, calculating variances of both sides of, we obtain:

$$\mathbb{V}\left(\sum_{i=1}^{n} X_i\right) = \mathbb{V}\,(b_n \cdot X)$$

$$n \cdot \mathbb{V}(X) = b_n^2 \cdot \mathbb{V}\,(X)$$

which, together with $\mathbb{V}(X) \neq 0$ (we exclude the degenerate case), immediately yields:

$$b_n = \sqrt{n}$$

that is the well-know results given by the Central Limit Theorem (with null first moment, leading to $a_n = n \cdot \mathbb{E}(X) = 0$).

We consider now the general case of summation of strictly stable random variables. Granted $S_n = \sum_{i=1}^{n} X_i$, rewriting the stability condition as the sequence of sums

$$S_1 = X_1 \overset{d}{\sim} b_1 \cdot X$$

$$S_2 = X_1 + X_2 \overset{d}{\sim} b_2 \cdot X$$

$$S_3 = X_1 + X_2 + X_3 \overset{d}{\sim} b_3 \cdot X$$

$$\dots$$

we consider only those sums which contain 2^k terms, with $k = 1, 2, ...$

$$S_2 = X_1 + X_2 \overset{d}{\sim} b_2 \cdot X$$

$$S_4 = X_1 + X_2 + X_3 + X_4 \overset{d}{\sim} b_4 \cdot X$$

$$S_8 = X_1 + X_2 + X_3 + X_4 + X_5 + X_6 + X_7 + X_8 \overset{d}{\sim} b_8 \cdot X$$

$$\dots$$

$$S_{2^k} = X_1 + X_2 + ... + X_{2^k - 1} + X_{2^k} \overset{d}{\sim} b_{2^k} \cdot X$$

$$\dots$$

Since the variables are identically distributed, we have $X_a + X_b \overset{d}{\sim} X_c + X_d$ for all $a, b, c, d \in \mathbb{N}$. That considered, we apply the relation involved in S_2 to S_4, that is:

$$S_4 = \underbrace{X_1 + X_2}_{X_1^*} + \underbrace{X_3 + X_4}_{X_2^*} \overset{d}{\sim} b_2 \cdot (X_1 + X_2) \overset{d}{\sim} b_2 \cdot (b_2 \cdot X) = b_2^2 \cdot X \,.$$

The same can be done in S_8:

$$\begin{aligned} S_8 = &\underbrace{X_1 + X_2}_{X_1^*} + \underbrace{X_3 + X_4}_{X_2^*} + \underbrace{X_5 + X_6}_{X_1^{**}} + \underbrace{X_7 + X_8}_{X_2^{**}} \overset{d}{\sim} \\ &\overset{d}{\sim} b_2 \cdot (X_1 + X_2) + b_2 \cdot (X_5 + X_6) \overset{d}{\sim} b_2^2 \cdot X_b + b_2^2 \cdot X_d = \\ &= b_2^2 \cdot (X_b + X_d) = b_2^2 \cdot (b_2 \cdot X) = b_2^3 \cdot X \,. \end{aligned}$$

Hence, for the sum of 2^k terms, we similarly obtain:

$$S_{2^k} \overset{d}{\sim} b_2^k \cdot X \,.$$

Comparing it with the expression of the definition of strictly stable random variable, with $n = 2^k$, we obtain:

$$b_n \cdot X \overset{d}{\sim} b_2^{\ln(n)/\ln(2)} \cdot X$$

that is

$$b_n = b_2^{\ln(n)/\ln(2)}$$

Rearranging this last expression[26], we get

$$b_n = n^{\frac{1}{\alpha_2}}$$

where $\alpha_2 = \ln(2)/\ln(b_2)$. Choosing now those sums which contain 3^k terms

$$S_{3^k} = X_1 + X_2 + ... + X_{3^k-1} + X_{3^k} \overset{d}{\sim} b_{3^k} \cdot X$$

and repeating the above reasoning, we arrive at

$$b_n = n^{\frac{1}{\alpha_3}}$$

where $\alpha_3 = \ln(3)/\ln(b_3)$. Thus, in the general case of sums with m^k terms

$$S_{m^k} = X_1 + X_2 + ... + X_{m^k-1} + X_{m^k} \overset{d}{\sim} b_{m^k} \cdot X$$

we have

$$b_n = n^{\frac{1}{\alpha_m}}$$

where $\alpha_m = \ln(m)/\ln(b_m)$.

Setting $m = 4$, according to the last expression, it ensues that

$$\alpha_4 = \frac{\ln(4)}{\ln(b_4)}\,.$$

Moreover, in the case of $n = 2^k$ addends, with $k = 2$, we have

$$b_4 = 4^{\frac{1}{\alpha_2}} = 4^{\ln(b_2)/\ln(2)}\,.$$

Comparing the two last formulae, we conclude that

$$\alpha_4 = \frac{\ln(4)}{\ln\left[4^{\ln(b_2)/\ln(2)}\right]} = \frac{\ln(2)}{\ln(b_2)} \cdot \frac{\ln(4)}{\ln(4)} = \frac{\ln(2)}{\ln(b_2)} = \alpha_2\,.$$

[26] It obvious by the properties of logarithms:

$$b_n = b_2^{\ln(n)/\ln(2)}$$

$$\ln(b_n) = \ln\left[b_2^{\ln(n)/\ln(2)}\right]$$

$$\ln(b_n) = \frac{\ln(n)}{\ln(2)} \cdot \ln(b_2)$$

$$\ln(b_n) = \ln\left[n^{\ln(b_2)/\ln(2)}\right]$$

$$b_n = n^{\ln(b_2)/\ln(2)}\,.$$

By induction, we come to the conclusion that, for all $m \in \mathbb{N}$, α_m are equal to each other:

$$\alpha_m = \alpha_{m-1} = \cdots = \alpha_3 = \alpha_2 = \alpha\,.$$

We call *index of stability* the common value α. The following expression hence holds for the scale factors b_n:

$$b_n = n^{\frac{1}{\alpha}}$$

whereas the original expression takes the form

$$\sum_{i=1}^{n} X_i \overset{d}{\sim} n^{\frac{1}{\alpha}} \cdot X\,.$$

□

A similar result can be achieved even for stable distribution (see Theorem 1.5.6). In any case, this theorem allows us to develop some useful observations. Substituting $b_n = n^{\frac{1}{\alpha}}$ into the expression obtained for the normal case, we have:

$$n \cdot \mathbb{V}(X) = b_n^2 \cdot \mathbb{V}(X) = n^{\frac{2}{\alpha}} \cdot \mathbb{V}(X)$$

So we see that for non-degenerate distributions with finite variance the index α must be equal to 2. If $\alpha \neq 2$, this relation can be "formally satisfied" only with $\mathbb{V}(X) = \infty$. Indeed, all stable distributions, except normal, have infinite variances, and some of them, such as the Cauchy's distribution, have also infinite mathematical expectation.

What follows exposes heuristically tails' behaviour of stable laws. As a matter of fact, our proposal is to show that there exists a connection between L-stable laws and the Pareto's distribution: This fact allow us to extend some of its properties, such has the infiniteness of the moments. Indeed, like Pareto's random variables, L-stable random variable have all infinite variance, except the Gaussian case. Moreover, L-stable variables have also infinite expected value if $\alpha \leq 1$. In order to give a qualitative explanation of that connection, we shall use some argumentations typical of Extreme Value Theory.

Let us consider a stable distribution with $\alpha < 1$. The arithmetic mean has the same distribution as

$$\frac{S_n}{n} = \frac{\sum_{i=1}^{n} X_i}{n} \overset{d}{\sim} \frac{n^{\frac{1}{\alpha}} \cdot X}{n} = n^{\frac{1-\alpha}{\alpha}} \cdot X$$

Meanwhile, the factor $n^{\frac{1-\alpha}{\alpha}}$ tends to infinity as n grows. Hence, we can say that the average of n variables X_i turns out considerably greater than any

fixed summand X_i. This is possible only in the case where the maximum term

$$M_n := \max\{X_1, X_2, \ldots, X_n\}$$

grows extremely fast and gives the greatest contribution to the sum S_n, as defined above. Using only simple arguments of rather heuristic nature, we can write the c.d.f. for the maximum M_n as:

$$\begin{aligned} F_{M_n}(x) &= \Pr\{M_n \leq x\} = \Pr\Big\{\max\{X_1, X_2, \ldots, X_n\} \leq x\Big\} = \\ &= \Pr\{X_1 \leq x, X_2 \leq x, \ldots, X_n \leq x\} \ . \end{aligned}$$

Since those variables are independent and identically distributed, we have:

$$\begin{aligned} F_{M_n}(x) &= \Pr\{X_1 \leq x\} \cdot \Pr\{X_2 \leq x\} \cdot \ldots \cdot \Pr\{X_n \leq x\} = \\ &= \Pr\{X \leq x\} \cdot \Pr\{X \leq x\} \cdot \ldots \cdot \Pr\{X \leq x\} = \\ &= (\Pr\{X \leq x\})^n = (1 - \Pr\{X > x\})^n \ . \end{aligned}$$

Furthermore, since $S_n \overset{d}{\sim} n^{\frac{1}{\alpha}} \cdot X$, we have:

$$\begin{aligned} F_{S_n}(x) &= \Pr\{S_n \leq x\} = \Pr\left\{n^{\frac{1}{\alpha}} \cdot X \leq x\right\} = \Pr\left\{X \leq n^{-\frac{1}{\alpha}} \cdot x\right\} = \\ &= 1 - \Pr\left\{X > n^{-\frac{1}{\alpha}} \cdot x\right\} \ . \end{aligned}$$

If the maximum gives a primary contribution to the sum S_n, the following relation should be satisfied:

$$F_{S_n}(x) \sim F_{M_n}(x)$$

when $x \to +\infty$. Since we are referring to *tail* behaviour, the previous relation can be written in a more convenient way[27]:

$$\Pr\{S_n > x\} \sim \Pr\{M_n > x\}$$

when $x \to +\infty$. Substituting the previous values, it follows[28]:

$$\Pr\{S_n > x\} \sim 1 - (1 - \Pr\{X > x\})^n \sim n \cdot \Pr\{X > x\} \ .$$

[27] Since $F_{S_n}(x) \sim F_{M_n}(x)$, it easily follows:

$$\begin{aligned} 1 - F_{S_n}(x) &\sim 1 - F_{M_n}(x) \\ \Pr\{S_n > x\} &\sim \Pr\{M_n > x\} \ . \end{aligned}$$

[28] To justify the last statement we have to prove that

$$\lim_{x \to +\infty} \frac{1 - (1 - \Pr\{X > x\})^n}{n \cdot \Pr\{X > x\}} = 1 \, .$$

As a matter of fact, as x increases we have $\Pr\{X > x\} \ll 1$. So, making the substitution $\Pr\{X > x\} = y$, the previous limit becomes:

$$\lim_{y \to 0^+} \frac{1 - (1 - y)^n}{n \cdot y} = 1 \, .$$

This relation (even named *tail preservation criterion*) expresses that the tail behaviour of X is preserved under finite addiction. Moreover, substituting even the left side, we arrive at:

$$\Pr\left\{X > n^{-\frac{1}{\alpha}} \cdot x\right\} \sim n \cdot \Pr\{X > x\} \,. \tag{1.33}$$

It can be proved that the solution of the previous equation is a Paretian random variable, thus having power-law tail distribution (see [40]):

$$\begin{aligned} \Pr\{X > x\} &\sim C_1 \cdot x^{-\alpha} \quad \text{if } x \to +\infty \\ \Pr\{X < x\} &\sim C_2 \cdot |x|^{-\alpha} \quad \text{if } x \to -\infty \end{aligned} \tag{1.34}$$

for some positive constant C_1 and C_2. Differentiating the last two expression with respect to x, we obtain the following asymptotic expressions for the density

$$\begin{aligned} \frac{\mathrm{d}\left(1 - \Pr\{X > x\}\right)}{\mathrm{d}x} &\sim \alpha \cdot C_1 \cdot x^{-\alpha-1} \quad \text{if } x \to +\infty \\ \frac{\mathrm{d}\left(\Pr\{X < x\}\right)}{\mathrm{d}x} &\sim \alpha \cdot C_2 \cdot |x|^{-\alpha-1} \quad \text{if } x \to -\infty \,. \end{aligned}$$

This results allows us to extend Pereto's distribution properties about tails (and moments, when feasible) to the L-stable ones. Hence (see [5] and [33]), if $k \in \mathbb{R}$ and $0 < \alpha < 2$, the absolute moments result (we can deal with fractional order moments as well)[29]:

$$\mathbb{E}\big(|X|^k\big) = \begin{cases} \infty & \text{if } 0 < \alpha \leq k \\ \text{finite} & \text{if } \alpha > k \,. \end{cases} \tag{1.35}$$

This rule is even valid for the mean, that is:

$$\mathbb{E}\big(X\big) = \begin{cases} \infty & \text{if } 0 < \alpha \leq 1 \\ \delta & \text{if } \alpha > 1 \,. \end{cases} \tag{1.36}$$

Thus, when $0 < \alpha < 2$, it ensues that $\mathbb{E}\big(|X|^2\big) = \mathbb{E}\big(X^2\big) = \infty$, and stable distributions do not have finite second moments or variances. However, since $\alpha = 2$ leads to the Gaussian law, we can synthesize the L-stable variables' mean and variance in the following table.

shape parameter	$\mathbb{E}(X)$	$\mathbb{V}(X)$
$0 < \alpha \leq 1$	∞	∞
$1 < \alpha < 2$	finite	∞
$\alpha = 2$	finite	finite

[29] The absolute value is required since the Pareto's distribution is defined only on a positive support.

Moreover the so-called skewness parameter β is not the same thing as the classical skewness parameter. The latter is undefined for every non-Gaussian stable distribution because neither the third moment or the variance exist. Likewise, the kurtosis is undefined, because the fourth moment is undefined for every non-Gaussian stable distribution.

We can state the following theorem, which generalize the previous result.

Theorem 1.5.5 *Given a L-stable random variable* $X \overset{d}{\sim} \mathcal{S}(\alpha, \beta, \gamma, \delta)$*, the following relations are valid*

$$\Pr\{X > x\} \sim \left[c(\alpha) \cdot (1+\beta) \cdot \gamma^{\alpha}\right] \cdot x^{-\alpha} \quad \text{if } x \to +\infty$$

$$\Pr\{X < x\} \sim \left[c(\alpha) \cdot (1-\beta) \cdot \gamma^{\alpha}\right] \cdot |x|^{-\alpha} \quad \text{if } x \to -\infty$$

where

$$c(\alpha) := \begin{cases} \frac{1}{2 \cdot \Gamma(1-\alpha) \cdot \cos\left(\alpha \cdot \frac{\pi}{2}\right)} & \text{if } \alpha \neq 1 \\ \frac{1}{\pi} & \text{if } \alpha = 1\,. \end{cases}$$

This theorem[30] implicitly says that, for all $\alpha < 2$ and $-1 < \beta < 1$, both tail probabilities are asymptotically power laws. When $\beta = -1$, the right tail of the distribution is not asymptotically a Pareto's law; likewise when $\beta = 1$, the left tail of the distribution is not asymptotically a power law.

The Generalized Central Limit Theorem ensues naturally.

Theorem 1.5.6 (Generalized Central Limit Theorem) *Let* $X_1, X_2, \ldots, X_n$ *be independent identically distributed random variables with tail probabilities that satisfy*

$$\lim_{x \to +\infty} \Pr\{X > x\} = C_1 \cdot x^{-k}$$

and

$$\lim_{x \to -\infty} \Pr\{X < x\} = C_2 \cdot |x|^{-k}$$

with $C_1 + C_2 > 0$ *and* $k > 0$*. Then there exist sequences* a_n *and* $b_n > 0$ *such that the distribution of the centered and normalized sum*

$$Z_n = \frac{\sum_{i=1}^{n} X_i - a_n}{b_n}$$

[30] In some textbook the function $c(\alpha)$ may be indicated as $c(\alpha) = \frac{\Gamma(\alpha) \cdot \sin\left(\alpha \cdot \frac{\pi}{2}\right)}{\pi}$. Reminding few properties of Gamma function, such as $\Gamma(z) \cdot \Gamma(1-z) = \frac{\pi}{\sin(\pi \cdot z)}$ and $\Gamma(1+z) = z \cdot \Gamma(z)$, and the double-angle formula for the sinus $\sin(2\theta) = 2 \cdot \sin(\theta) \cdot \cos(\theta)$, the prove they are the same quantity is quite trivial.

converges in distribution to the stable distribution $X \overset{d}{\sim} \mathcal{S}(\alpha, \beta, 1, 0)$ *with parameters*

$$\alpha := \begin{cases} k & \text{if } k < 2 \\ 2 & \text{if } k \geq 2 \end{cases} \qquad \textit{and} \qquad \beta := \frac{C_1 - C_2}{C_1 + C_2}\,.$$

Moreover, the coefficients a_n *and* b_n *are those written in the following table.*

k	α	a_n	b_n
$0 < k < 1$	k	0	$\left[\Gamma(1-\alpha)\cdot\cos\left(\alpha\cdot\frac{\pi}{2}\right)\cdot(C_1+C_2)\right]^{\frac{1}{\alpha}}\cdot n^{\frac{1}{\alpha}}$
$k = 1$	1	$(C_1 - C_2)\cdot\ln(n^n)$	$\frac{\pi}{2}\cdot(C_1+C_2)\cdot n$
$1 < k < 2$	k	$n\cdot\mathbb{E}(X)$	$\left[\Gamma(1-\alpha)\cdot\cos\left(\alpha\cdot\frac{\pi}{2}\right)\cdot(C_1+C_2)\right]^{\frac{1}{\alpha}}\cdot n^{\frac{1}{\alpha}}$
$k \geq 2$	2	$n\cdot\mathbb{E}(X)$	$\sqrt{\frac{\mathbb{V}(X)}{2}\cdot n}$

Considering in more detail the summation of non-negative random variables X_i which satisfy the condition of Theorem 1.5.6, we arrive at the following qualitative conclusions.

For $0 < \alpha < 1$, the "typical" values of the sum $S_n = \sum_{i=1}^n X_i$ (for example, the most probable value corresponding to the position of the maximum of the density) behave as $n^{\frac{1}{\alpha}}$, that is, it increase much faster than in the case where $\alpha > 1$ since, in that case, the expectation exists. A "dispersion" of values of the random sum S_n grows with the same rate, so its relative fluctuations (the ratio of the width to the most probable value) do not vanish.

For $1 < \alpha < 2$, there exists $\mathbb{E}(X)$, and in view of the Law of Large Numbers the typical value of the sum is proportional to the number of summands. However, the "dissipation" of the sum characterized by the factor b_n grows as $n^{\frac{1}{\alpha}}$. Thus, the relative fluctuations decrease as $n^{\frac{1-\alpha}{\alpha}}$, that is, essentially slower than in the normal case, where finiteness of the variance leads us to the traditional behaviour of the relative fluctuations which grow as $n^{-\frac{1}{2}}$.

Generalized Central Limit Theorem and the conclusion we have made help us in giving a quantitative definition of *heavy tailed distributions.*

Definition 1.5.7 *Given a random variable* X*, we say it has* heavy tails *if it satisfies the following conditions:*

$$\Pr\{X > x\} = \int_x^{+\infty} f_X(x)\,\mathrm{d}x \sim C_1\cdot x^{-\alpha} \qquad \text{if } x \to +\infty$$

$$\Pr\{X < x\} = \int_{-\infty}^{x} f_X(x)\,\mathrm{d}x \sim C_2\cdot |x|^{-\alpha} \qquad \text{if } x \to -\infty\,.$$

On the contrary we say it has short tails *if*

$$\Pr\{X > x\} = \int_x^{+\infty} f_X(x)\,\mathrm{d}x = o\left(x^{-2}\right) \qquad \text{if } x \to +\infty$$

$$\Pr\{X < x\} = \int_{-\infty}^{x} f_X(x)\,\mathrm{d}x = o\left(x^{-2}\right) \qquad \text{if } x \to -\infty\,.$$

Finally we only state some fundamental properties of stable distributions. The most useful and crucial one is that sums of L-stable random variables are L-stable. In the independent case, the exact parameters of the sums are given below. When the summands are dependent, the sum is stable but the precise statement is more difficult and depends on the exact dependence structure.

Proposition 1.5.8 *Suppose that $X_1, X_2, ..., X_n$ are independent and identically distributed random variables with $X_i \overset{d}{\sim} \mathcal{S}(\alpha, \beta_i, \gamma_i, \delta_i)$ and $i = 1, 2, ..., n$. Then the following relations hold true*:

- *The distribution of*

$$Y := \sum_{i=1}^{n} X_i$$

is L-stable, moreover

$$Y \overset{d}{\sim} \mathcal{S}\left[\alpha\,,\ \frac{\sum_{i=1}^{n} \beta_i \cdot \gamma_i^{\alpha}}{\sum_{i=1}^{n} \gamma_i^{\alpha}}\,,\ \left(\sum_{i=1}^{n} \gamma_i^{\alpha}\right)^{\frac{1}{\alpha}}\,,\ \sum_{i=1}^{n} \delta_i\right]\,; \tag{1.37}$$

- *The distribution of*

$$Y := h \cdot X_i + k$$

for some real constant $h \neq 0$ and k, is L-stable such that

$$Y \overset{d}{\sim} \mathcal{S}\left[\alpha\,,\ \mathrm{sgn}(h) \cdot \beta_i\,,\ |h| \cdot \gamma_i\,,\ \delta^*\right] \tag{1.38}$$

where

$$\delta^* := \begin{cases} h \cdot \delta_i + k & \text{if } \alpha \neq 1 \\ h \cdot \delta_i - \frac{2}{\pi} \cdot \ln\left(|h|^h\right) \cdot \gamma_i \cdot \beta_i + k & \text{if } \alpha = 1\,; \end{cases}$$

- *For all X_i, the characteristic functions, densities and distribution functions are continuous away from $\alpha = 1$, but discontinuous in any neighborhood of $\alpha = 1$.*

The proposition proves that the shape of the sum of n terms is the same as the original shape. We stress that *no other* class of distributions owns this property (by virtue of that, it is said that the L-stable distributions represent

the *domain of attraction* for *every* sum of i.i.d. random variables, even if they exhibits infinite variance).

Furthermore, it can be shown that all stable distributions with $\alpha < 1$ are strictly stable, and the stable distributions with $\alpha > 1$ are easily transformable to strictly stable ones since $\mathbb{E}(X)$ exists and is used for centering the random variable. The distribution with $\alpha = 1$ is strictly stable if only $\beta = 0$ (Cauchy's distribution).

Without entering into details, the reason we introduced such distribution is involved in features of financial data. As a matter of fact, they present volatility clustering, which can be seen as a consequence of divergence of the variance. We have seen that L-stable random variable with $\alpha < 2$ possesses that property.

Moreover, we shall see later, the Mandelbrot's fractal concept has found its application in the turbulence problem, anomalous diffusion, dynamical chaos and large-scale structure of phenomena. In the foundation of these topics, the main part is taken by the power-type laws that provide us with the *scale invariance* (self-affinity) of the structures and processes. Thus, the stable laws with their power-type asymptotic behavior and infinite variance turn out to be very suitable.

1.6 Lévy Processes (LP) and Lévy-stable motion (LSM)

The last section of this chapter is devoted to the introduction and analysis of Lévy-stable process (LSP). They form a specific subclass of *Lévy processes* (LP). Moreover we shall see that LP and LSP are alternative generalizations of the standard Brownian motion. They provide a richer and more versatile environment to model the behaviour of financial markets than those purely based on SBM. As a matter of fact, LSM shares unique characteristics to their class such as *stability* (linear combinations of different random variables have again a L-stable distribution, thus having the same shape as the individual random variables up to scale and shift) and they can easily accommodate *heavy tails* and *skewness* of stock returns, a much desired property in empirical finance.

Let we start introducing Lèvy processes (LP).

Definition 1.6.1 *Given a probability space* $(\Omega, \mathcal{F}, \mathbf{P})$*, every real stochastic process* $\{L(t)\}_{t \in [0,+\infty)}$ *defined on it is said to be a* Lévy process*, if it possesses the following properties:*

(a) $L(0) = 0$ *almost surely;*

(b) *for every* $0 = t_0 < t_1 < t_2 < ... < t_n < +\infty$ *all the increments* $L(t_1) - L(t_0)\,,\, L(t_2) - L(t_1)\,,\, ...\,,\, L(t_{n-1}) - L(t_n)$ *are independent;*

(c) *the increments are stationary, that is* $L(t)-L(s) \overset{d}{\sim} L(t+\tau)-L(s+\tau)$, *for all* $s<t, \tau \in T$;

(d) $\{L(t)\}_{t\in[0,+\infty)}$ *is a càdlag process almost surely.*

First of all we can see that, if we compare it to a SBM (see Definition 1.1.1 and 1.1.10), we made only few condition less demanding. This allows us to state that the standard Brownian motion is a special case of LP. Hence all the next results will be applicable to the Brownian motion as well.

However, from Definition 4.9 alone it is difficult to see how rich the class of Lévy processes is. The Italian mathematician Bruno De Finetti introduced the notion of an *infinitely divisible* distribution and showed that they have an intimate relationship with Lévy processes (and stability as well).

Definition 1.6.2 *The law* F_X *of a random variable* X *is* infinitely divisible, *if there exist a sequence of i.i.d. random variables* $X_1^{(1/n)}, X_2^{(1/n)}, ..., X_n^{(1/n)}$ *such that, for all* $n \in \mathbb{N}$

$$X \overset{d}{\sim} \sum_{i=1}^{n} X_i^{(1/n)}. \tag{1.39}$$

Equivalently, the law F_X *of a random variable* X *is infinitely divisible if for all* $n \in \mathbb{N}$ *there exists another law* $F_{X^{(1/n)}}$ *of a random variable* $X^{(1/n)}$ *such that*

$$F_X = \underbrace{F_{X^{(1/n)}} * ... * F_{X^{(1/n)}}}_{n \text{ times}}.$$

This means that the sequence of i.i.d. random variable belong to the same family of distribution of X but does not own its same finite-dimensional properties (for instance, they have different moments). Alternatively, we can characterize an infinitely divisible random variable using its characteristic function.

Proposition 1.6.3 *The law of a random variable* X *is infinitely divisible, if for all* $n \in \mathbb{N}$, *there exists a random variable* $X^{(1/n)}$, *such that*

$$\varphi_X(u) = \left[\varphi_{X^{(1/n)}}(u)\right]^n. \tag{1.40}$$

The logarithm of the characteristic function is called *characteristic exponent* and it is denoted by $\Psi_X(u)$. Hence the previous expression allows us to write:

$$\Psi_X(u) = \ln\left[\varphi_X(u)\right] = \ln\left\{\left[\varphi_{X^{(1/n)}}(u)\right]^n\right\} = n\cdot\ln\left[\varphi_{X^{(1/n)}}(u)\right] = n\cdot\Psi_{X^{(1/n)}}(u)$$

The simplest example of infinitely divisible distribution is the normal law. Suppose $X \overset{d}{\sim} \mathcal{N}\left(\mu, \sigma^2\right)$., then we have:

$$\varphi_X(u) = e^{\imath\,\mu\,u - \frac{\sigma^2\,u^2}{2}} = e^{n\cdot\left(\frac{\imath\,\mu\,u}{n} - \frac{\sigma^2\,u^2}{2\,n}\right)} = e^{\left(\imath\,u\cdot\frac{\mu}{n} - \frac{u^2}{2}\cdot\frac{\sigma^2}{n}\right)^n}.$$

Setting $\mu_{X^{(1/n)}} = \frac{\mu}{n}$ and $\sigma^2_{X^{(1/n)}} = \frac{\sigma^2}{n}$, we have:

$$\varphi_X(u) = e^{\left(\imath\, u \cdot \mu_{X^{(1/n)}} - \frac{u^2}{2} \cdot \sigma^2_{X^{(1/n)}}\right)^n} = \left[\varphi_{X^{(1/n)}}(u)\right]^n .$$

Thus the sequences of i.i.d random variables which satisfy the infinitely divisible condition is given by $X_i^{(1/n)} \overset{d}{\sim} \mathcal{N}\left(\frac{\mu}{n}, \frac{\sigma^2}{n}\right)$.

The next theorem provides a complete characterization of random variables with infinitely divisible distributions via their characteristic functions; this is the celebrated *Lévy-Khintchine formula*. Its proof can be found in [39].

Theorem 1.6.4 (**Lévy-Khintchine formula**)*A probability law F_X of a real-valued random variable X is infinitely divisible if and only if there exists a triple $(a, b, \mathbf{\Lambda})$, where $a \in \mathbb{R}$, $b \in \mathbb{R}_0^+$ and $\mathbf{\Lambda}$ is a measure concentrated on $\mathbb{R}\backslash\{0\}$ satisfying $\int_{\mathbb{R}} \min\left\{1, x^2\right\} \mathbf{\Lambda}(\mathrm{d}x) < +\infty$ such that its characteristic exponent is*

$$\Psi_X(u) = \imath\, a\, u - \frac{b^2\, u^2}{2} + \int_{|x|\geq 1} (1 - e^{\imath\, u\, x})\ \mathbf{\Lambda}(\mathrm{d}x) + \int_{0<|x|<1} (1 - e^{\imath\, u\, x} + \imath\, u\, x)\ \mathbf{\Lambda}(\mathrm{d}x) .$$

The triplet $(a, b, \mathbf{\Lambda})$ is named Lévy (or characteristic) triplet. Moreover, a is called the *drift term*, b the Gaussian or *diffusion coefficient* and the measure $\mathbf{\Lambda}$ the *Lévy measure*. It is important to note that the latter is unique. Let us now discuss in further detail the relationship between infinitely divisible distributions and processes with stationary independent increments.

Consider a Lévy process $\{L(t)\}_{t\in[0,+\infty)}$; for any $n \in \mathbb{N}$ we trivially have that

$$\begin{aligned} L(t) &= L\left(t \cdot \frac{1}{n}\right) + \left[L\left(t \cdot \frac{2}{n}\right) - L\left(t \cdot \frac{1}{n}\right)\right] + ... + \left[L\left(t\right) - L\left(t \cdot \frac{n-1}{n}\right)\right] \\ &= \sum_{k=1}^{n} L\left(t \cdot \frac{k}{n}\right) - L\left(t \cdot \frac{k-1}{n}\right) . \end{aligned}$$

The stationarity and independence of the increments yield that

$$\left\{L\left(t \cdot \frac{k}{n}\right) - L\left(t \cdot \frac{k-1}{n}\right)\right\}_{k\in[1,n]}$$

is an i.i.d. sequence of random variables; hence we can conclude that the random variable $L(t)$ is infinitely divisible. It follows the next theorem.

Theorem 1.6.5 *For every Lévy process $\{L(t)\}_{t\in[0,+\infty)}$ we have that*

$$\varphi_{L(t)}(u) = e^{t\,\cdot\,\Psi_{L(1)}(u)}$$

where $\varphi_{L(t)}(u)$ is the characteristic function of the process and $\Psi_{L(1)}(u)$ is the characteristic exponent of $L(1)$, a random variable with an infinitely divisible distribution.

Proof: Since

$$\varphi_{L(t+s)}(u) = \mathbb{E}\left[e^{\imath\cdot u\cdot L(t+s)}\right] = \mathbb{E}\left\{e^{\imath\cdot u\cdot\left[L(t+s)-L(s)+L(s)\right]}\right\} =$$
$$= \mathbb{E}\left\{e^{\imath\cdot u\cdot\left[L(t+s)-L(s)\right]}\cdot e^{\imath\cdot u\cdot L(s)}\right\}$$

due to the independence and the stationarity of the increments, we have:

$$\varphi_{L(t+s)}(u) = \mathbb{E}\left\{e^{\imath\cdot u\cdot\left[L(t+s)-L(s)\right]}\right\}\cdot\mathbb{E}\left[e^{\imath\cdot u\cdot L(s)}\right] =$$
$$= \mathbb{E}\left[e^{\imath\cdot u\cdot L(t)}\right]\cdot\mathbb{E}\left[e^{\imath\cdot u\cdot L(s)}\right] = \varphi_{L(t)}(u)\cdot\varphi_{L(s)}(u)\,.$$

Now we have the boundary condition $\varphi_{L(0)}(u) = 1$; the unique continuous solution of the previous Cauchy functional equation is

$$\varphi_{L(t)}(u) = e^{t\cdot f(u)}$$

where $f:\mathbb{R}\to\mathbb{C}$. Since $L(1)$ is an infinitely divisible random variable, the statement $f(u) = \Psi_{L(1)}(u)$ follows.

□

Moreover, by virtue of Lévy-Khintchine formula, it easy to prove that even arithmetical Brownian motion (see Definition 1.1.13) is a Lévy process (we have already shown that the SBM is a Lévy process). Given the normal law with

$$\varphi_X(u) = e^{\imath\,\mu\,u-\frac{\sigma^2\,u^2}{2}}$$

that is $\Psi_X(u) = \imath\,\mu\,u-\frac{\sigma^2\,u^2}{2}$, if we set the Lévy triplet as $(a=\mu, b=\sigma, \mathbf{\Lambda}=0)$, by Theorem 1.6.4, we get

$$\Psi_X(u) = \imath\,\mu\,u - \frac{\sigma^2\,u^2}{2} + \int_{|x|\geq 1}(1-e^{\imath\,u\,x})\cdot 0 + \int_{0<|x|<1}(1-e^{\imath\,u\,x}+\imath\,u\,x)\cdot 0 =$$
$$= \imath\,\mu\,u - \frac{\sigma^2\,u^2}{2}\,.$$

Hence, given a SBM $\{B(t)\}_{t\,\in\,[0,+\infty)}$ the process

$$B^*(t) = \mu\cdot t + \sigma\cdot B(t)$$

it has characteristic function

$$\varphi_{B^*(t)}(u) = \mathbb{E}\left[e^{\imath\,u\,B^*(t)}\right] = \mathbb{E}\left\{e^{\imath\,u\,[\mu\,t+\sigma\,B(t)]}\right\} = \mathbb{E}\left[e^{\imath\,u\,\mu\,t+\imath\,u\,\sigma\,B(t)}\right] =$$
$$= \mathbb{E}\left[e^{\imath\,u\,\mu\,t}\cdot e^{\imath\,u\,\sigma\,B(t)}\right] = e^{\imath\,u\,\mu\,t}\cdot\mathbb{E}\left[e^{\imath\,u\,\sigma\,B(t)}\right]\,.$$

Since $B(t) \overset{d}{\sim} \mathcal{N}(0,t)$, it follows that $\sigma \cdot B(t) \overset{d}{\sim} \mathcal{N}(0,\sigma^2\, t)$. Moreover, the last expected value is nothing but the characteristic function of the scaled random variable $\sigma \cdot B(t)$. Hence:

$$\varphi_{B^*(t)}(u) = e^{\imath\, u\, \mu\, t} \cdot e^{\imath \cdot 0 \cdot u - \frac{(\sigma^2\, t)\, u^2}{2}} = e^{t\,(\imath\, u\, \mu)} \cdot e^{t\left(-\frac{\sigma^2\, u^2}{2}\right)} =$$
$$= e^{t\left(\imath\, u\, \mu - \frac{\sigma^2\, u^2}{2}\right)} = e^{t \cdot \Psi_{B^*(1)}(u)}$$

where $B^*(1) \overset{d}{\sim} \mathcal{N}(\mu, \sigma^2)$. By Theorem 1.6.5, this proves that BM is a Lévy process as well.

Now we will only state a crucial result about Lévy process. We will not prove it (see [39]) but only examine few of its implication, without a in-depth discussion. Moreover, it will be very useful in describing the paths' variations of LP.

Theorem 1.6.6 (**Lévy-Itō decomposition**) *Let $\{L(t)\}_{t\,\in\,[0,+\infty)}$ be a Lévy process characterized by the Lévy triplet $(a, b, \mathbf{\Lambda})$. If the condition*

$$\int_{\mathbb{R}} \min\left\{1, x^2\right\}\, \mathbf{\Lambda}(\mathrm{d}x) < \infty$$

holds, then there is a probability space $(\Omega, \mathcal{F}, \mathbf{P})$ on which three independent Lévy processes exist, $L^{(1)}(t)$, $L^{(2)}(t)$, $L^{(3)}(t)$, where:

- *$L^{(1)}(t)$ is a Brownian motion;*
- *$L^{(2)}(t)$ is a compound Poisson process*[31]*;*
- *$L^{(3)}(t)$ is a square-integrable martingale*[32]*;*

such that the original Lévy process can be written as:

$$L(t) = L^{(1)}(t) + L^{(2)}(t) + L^{(3)}(t) \tag{1.41}$$

[31] A *Poisson process* $\{N(t)\}_{t\,\in\,[0,+\infty)}$ is a stochastic process which counts the number of events and the time occurred between them in a given time interval. The time between each pair of consecutive events has an exponential distribution with parameter $\lambda > 0$ and each of these inter-arrival times is assumed to be independent of other inter-arrival times. A *compound Poisson process* $\{P(t)\}_{t\,\in\,[0,+\infty)}$ with intensity λ is defined by

$$P(t) = \sum_{i=i}^{N(t)} A_i$$

where A_i are independent and identically distributed random variables, which are also independent of $\{N(t)\}$. Both are Lévy processes.

[32] A *square-integrable martingale* $\{X(t)\}_{t\,\in\,[0,+\infty)}$ is a martingale such that $\mathbb{E}\left[X(t)^2\right]$ exists and is finite.

Hence, Lévy–Itō decomposition describes the structure of a general Lévy process in terms of three independent auxiliary Lévy processes, each with different types of path behaviour. Note that $\{L^{(1)}(t)\}$ is a continuous process, whereas the other two processes are discontinuous. Thus, heuristically, we can say LP can have both continuous (if the last two processes are null a.s.) or discontinuous (otherwise) paths.

According to Theorem 1.6.4, any characteristic exponent belonging to an infinitely divisible distribution can be written, after some simple reorganisation, in the form

$$\Psi_L(u) = \Psi_L^{(1)}(u) + \Psi_L^{(2)}(u) + \Psi_L^{(3)}(u)$$

where

$$\begin{aligned}
\Psi_L^{(1)}(u) &= \imath\, a\, u - \frac{b^2\, u^2}{2} \\
\Psi_L^{(2)}(u) &= \int_{|x|\geq 1} (1 - e^{\imath\, u\, x})\ \mathbf{\Lambda}(\mathrm{d}x) \\
\Psi_L^{(3)}(u) &= \int_{0<|x|<1} (1 - e^{\imath\, u\, x} + \imath\, u\, x)\ \mathbf{\Lambda}(\mathrm{d}x)\,.
\end{aligned}$$

The first one, we have already seen, is the characteristic exponent of an arithmetical Brownian motion. It is possible to prove (see [39]) that the second one is the characteristic exponent of a compound Poisson process with intensity $\lambda = \mathbf{\Lambda}(\mathbb{R}\backslash(-1,1))$. The last part is the most difficult to handle, but it correspond to the overlapping of (at most) a countable number of independent compound Poisson processes with different arrival rates and additional linear drift.

Furthermore, by virtue of Definition A.4.5 and Lévy–Itō decomposition every Lévy process is also a *semimartingale*.

The last part of this overall introduction to Lévy processes is dedicated to the description of their path variations. More specifically, the Lévy measure is responsible for the multiplicity of the class of Lévy processes and carries useful information about the structure of the process. The process can have a finite number of jumps on every time interval (which is addressed by *finite activity*) or an infinite on (conversely *infinite activity*). We will state few theorems about LP activity. Their proof can be found in [39]. As previously said, path properties have to be ascribed to the Lévy measure.

Proposition 1.6.7 *: Let $\{L(t)\}_{t\in[0,+\infty)}$ be a Lévy process with characteristic triplet $(a, b, \mathbf{\Lambda})$.*

(a) *If $\mathbf{\Lambda}(\mathbb{R}) < \infty$, then almost all paths of $\{L(t)\}$ have a finite number of jumps on every compact interval. In that case, the Lévy process has* finite activity.

(b) *If* $\mathbf{\Lambda}(\mathbb{R}) = \infty$*, then almost all paths of* $\{L(t)\}$ *have an infinite number of jumps on every compact interval. In that case, the Lévy process has* infinite activity.

Whether a Lévy process has finite variation or not also depends on the Lévy measure (and on the presence/absence of a Brownian part).

Proposition 1.6.8 *: Let* $\{L(t)\}_{t\in[0,+\infty)}$ *be a Lévy process with characteristic triplet* $(a, b, \mathbf{\Lambda})$.

(a) *If* $b = 0$ *and* $\int_{|x|\geq 1}|x|\,\mathbf{\Lambda}(\mathrm{d}x) < \infty$ *then almost all paths of* $\{L(t)\}$ *have a* finite variation.

(b) *If* $b \neq 0$ *or* $\int_{|x|\geq 1}|x|\,\mathbf{\Lambda}(\mathrm{d}x) = \infty$ *then almost all paths of* $\{L(t)\}$ *have a* infinite variation.

The Lévy measure also carries information about the finiteness of the moments of a Lévy process. This is particularly useful information in mathematical finance, related to the existence of a martingale measure. The finiteness of the moments of a LP is related to the finiteness of an integral over the Lévy measure.

Proposition 1.6.9 *: Let* $\{L(t)\}_{t\in[0,+\infty)}$ *be a Lévy process with characteristic triplet* $(a, b, \mathbf{\Lambda})$*. It has finite p-th moment, that is* $\mathbb{E}\big[|L(t)|^p\big] < \infty$*, if and only if* $\int_{|x|\geq 1}|x|^p\,\mathbf{\Lambda}(\mathrm{d}x) < \infty$.

Eventually, we will focus on a particular subclass of LP: Lévy-stable processes (LSP). We will show that by choosing a particular triplet $(a, b, \mathbf{\Lambda})$ we are able to obtain it. Moreover we shall see that a subcase of Lévy-stable processes is represented by the Lèvy-stable motion (LSM).

Stable processes are the class of Lévy processes whose characteristic exponents correspond to those of L-stable distributions. Stable distributions were introduced by Lévy as a third example of infinitely divisible distributions after Gaussian and Poisson distributions. Recalling that a random variable X is L-stable if there exist constants a_n and $b_n > 0$ such that

$$\sum_{i=1}^{n} X_i \overset{d}{\sim} a_n + b_n \cdot X$$

where X_1, X_2, ..., X_n are independent random variables each having the same distribution as $X \overset{d}{\sim} \mathcal{S}(\alpha, \beta, \gamma, \delta)$, by subtracting $\frac{a_n}{n}$ from each of the terms on the left-hand side of the expression above and dividing by b_n, one sees in particular that this definition implies that any stable random variable is infinitely divisible. That is

$$\sum_{i=1}^{n} X_i - n \cdot \frac{a_n}{n} \overset{d}{\sim} b_n \cdot X$$

$$\sum_{i=1}^{n}\left(X_i - \frac{a_n}{n}\right) \overset{d}{\sim} b_n \cdot X$$

$$\sum_{i=1}^{n}\frac{X_i - \frac{a_n}{n}}{b_n} \overset{d}{\sim} X\,. \tag{1.42}$$

Setting $X_i^{(1/n)} = \frac{X_i - \frac{a_n}{n}}{b_n}$, it coincides with the infinitely divisible definition. Applying the same technique we used for the normal case, the corresponding sequence of i.i.d. random variable is given by:

$$X_i^{(1/n)} \overset{d}{\sim} \mathcal{S}\left(\alpha, \beta, \frac{\gamma}{\sqrt[\alpha]{n}}, \frac{\delta}{n}\right)\,. \tag{1.43}$$

For an alternative prove, apply Proposition 1.5.8 to $X_i^{(1/n)}$.

Since the characteristic function is known, we have the characteristic exponent of a L-stable random variable equal to

$$\Psi_X(u) = \begin{cases} \imath\delta u - \gamma^\alpha |u|^\alpha \cdot \left[1 + \imath\beta \cdot \operatorname{sgn}(u) \cdot \tan\left(\alpha \cdot \frac{\pi}{2}\right)\right] & \text{if } \alpha \neq 1 \\ \imath\delta u - \gamma|u| \cdot \left[1 + \frac{2\imath\beta}{\pi} \cdot \operatorname{sgn}(u) \cdot \ln|u|\right] & \text{if } \alpha = 1 \end{cases} \tag{1.44}$$

with the usual parameter restrictions.

Definition 1.6.10 *Given a probability space $(\Omega, \mathcal{F}, \mathbf{P})$ and a Lévy process $\{M^*(t)\}_{t\in[0,+\infty)}$ defined on it, $\{M^*(t)\}_{t\in[0,+\infty)}$ is said to be a* L-stable process *if all its finite-dimensional distributions are L-Stable.*

Thanks to Theorem 1.6.5, the characteristic function of the process is given by

$$\varphi_{M^*(t)}(u) = e^{t\cdot\Psi_{M^*(1)}(u)}$$

where $M^*(1)$ is a L-stable random variable such that $M^*(1) \overset{d}{\sim} \mathcal{S}(\alpha, \beta, \gamma, \delta)$ and its characteristic exponent $\Psi_{M^*(1)}(u)$ is equal to (1.44). Thanks to Proposition 1.5.8 and to the definition of infinitely divisibility, it follows, for $t \geq 0$

$$M^*(t) \overset{d}{\sim} \mathcal{S}\left(\alpha\,,\,\beta\,,\,t^{\frac{1}{\alpha}}\cdot\gamma\,,\,t\cdot\delta\right)\,. \tag{1.45}$$

Moreover, for $s, t \geq 0$,

$$M^*(t) - M^*(s) \overset{d}{\sim} \mathcal{S}\left(\alpha\,,\,\beta\,,\,|t-s|^{\frac{1}{\alpha}}\cdot\gamma\,,\,|t-s|\cdot\delta\right)\,. \tag{1.46}$$

To make the connection with the Lévy–Khintchine formula (we ignore the Brownian motion case, that is $\alpha = 2$) we need a Lèvy triplet $(a, b, \mathbf{\Lambda})$ such that $a = \delta$, $b = 0$ and

$$\mathbf{\Lambda}(\mathrm{d}x) = \begin{cases} C_1 \cdot x^{-\alpha-1}\,\mathrm{d}x & \text{if } x \in \mathbb{R}^+ \\ C_2 \cdot |x|^{-\alpha-1}\,\mathrm{d}x & \text{if } x \in \mathbb{R}^- \end{cases}$$

where $C_1, C_2 > 0$ and, if $\alpha \in (0,1) \cup (1,2)$,

$$\beta = \frac{C_1 - C_2}{C_1 + C_2} \qquad \text{and} \qquad \gamma = \sqrt[\alpha]{C_1^\alpha + C_2^\alpha}$$

or $C_1 = C_2$, if $\alpha = 1$. Straightforward integration yields the result.

Considering Lévy-Itō decomposition, LSP are composed by a deterministic linear process (drift) with parameter δ, a compound Poisson process and a square integrable martingale. This allows us to say that there are infinitely many jumps in any finite time interval, which means the paths are purely discontinuous. As a matter of fact, it is possible to prove that

$$\mathbf{\Lambda}(\mathbb{R}) = \int_{\mathbb{R}} \mathbf{\Lambda}(\mathrm{d}x) = \infty$$

and hence the L-stable process is a *infinite activity process*. On the other hand, since we have

$$\int_{|x|\geq 1} |x| \, \mathbf{\Lambda}(\mathrm{d}x) = \begin{cases} \frac{C_1+C_2}{1-\alpha} & \text{if } \alpha < 1 \\ \infty & \text{if } \alpha \geq 1 \end{cases}$$

we conclude that LSP has finite variation if $\alpha < 1$ and infinite variation if $\alpha \geq 1$.

The last notion of this section is represented by Lévy stable motion (LSM), which is a LSP with null location parameter. We have the following definition.

Definition 1.6.11 *Given a probability space $(\Omega, \mathcal{F}, \mathbf{P})$ and a L-stable process $\{M(t)\}_{t\in[0,+\infty)}$ defined on it, $\{M(t)\}_{t\in[0,+\infty)}$ is a* Lévy stable motion *if, for every $0 \leq s < t$, all increments $M(t) - M(s)$ are zero-centred stationary L-stable random variable such that*

$$M(t) - M(s) \overset{d}{\sim} \mathcal{S}\left(\alpha, \beta, \sqrt[\alpha]{t-s} \cdot \gamma, 0\right) . \tag{1.47}$$

If in addition $\gamma = 1$, we have a standard LSM. Of course, if $\alpha = 2$, we have a standard Brownian motion. Furthermore the following result ensues trivially.

Theorem 1.6.12 *Given a probability space $(\Omega, \mathcal{F}, \mathbf{P})$, the L-stable motion $\{M(t)\}_{t\in[0,+\infty)}$ defined on it is a self-affine process with Hurst index equal to $\frac{1}{\alpha}$ with $\alpha \in (1,2]$.*

Proof: Since $M(0) = 0$ a.s. and by the definition, given $a > 0$ we have

$$M(a \cdot t) \overset{d}{\sim} \mathcal{S}\left[\alpha\,,\, \beta\,,\, (a \cdot t)^{\frac{1}{\alpha}} \cdot \gamma\,,\, 0\right] .$$

Moreover, since $\alpha \in (0, 2]$, it follows $a^{\frac{1}{\alpha}} > 0$. By virtue of Proposition 1.5.8, the distribution scaled by $a^{\frac{1}{\alpha}}$ is such that

$$a^{\frac{1}{\alpha}} \cdot M(t) \overset{d}{\sim} \mathcal{S}\left[\alpha\,,\,\beta\,,\,(a \cdot t)^{\frac{1}{\alpha}} \cdot \gamma\,,\,0\right]\,.$$

However note that, since by point 6 of Proposition A.3.2, the Hurst exponent must be $0 < H < 1$. Hence, in the LSM case we have these two conditions to take into account, that is

$$\begin{cases} 0 < \alpha \leq 2 \\ 0 < \frac{1}{\alpha} < 1 \end{cases} \rightarrow \begin{cases} 0 < \alpha \leq 2 \\ \{\alpha > 0\} \cap \{\alpha < 0 \cup \alpha > 1\} \end{cases} \rightarrow \begin{cases} 0 < \alpha \leq 2 \\ \alpha > 1\,. \end{cases}$$

So in order to be *both* stable *and* self-affine of Hurst exponent $H = \frac{1}{\alpha}$, the condition

$$1 < \alpha \leq 2$$

must hold.

□

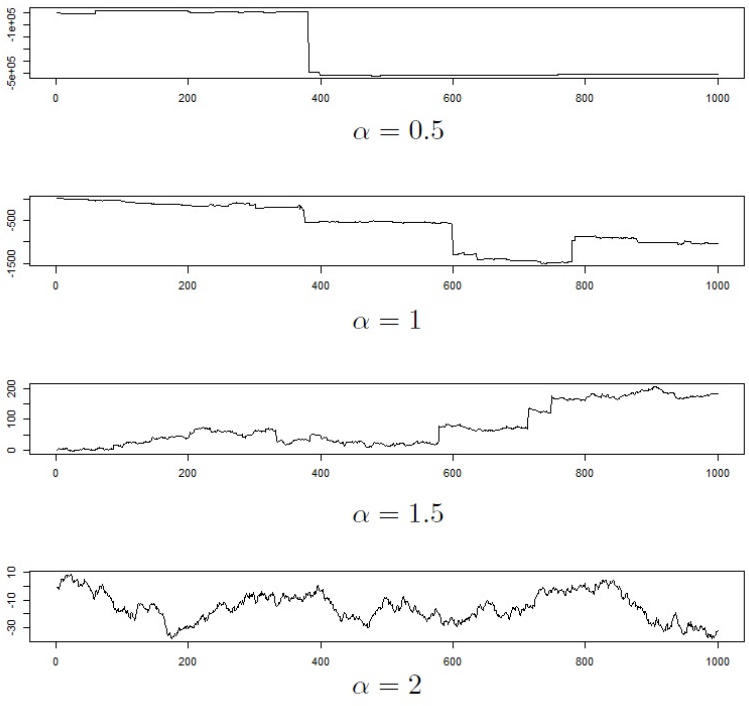


Figure 1.4: Discrete versions of symmetric standard Lévy stable motions

It is interesting to note that, if $1 < \alpha \leq 2$, it also implies

$$\frac{1}{2} \leq \frac{1}{\alpha} < 1$$

hence, the Hurst exponent is the same of a FBM with long-range dependence (but see below for an extensive discussion).

Being $1 < \alpha \leq 2$, on one hand the single random variables $M(t)$ need al least a finite first moment and, on the other hand the process itself exhibits infinite variation. Given such a process, it is possible to compute its covariance function. Since it has finite mean $\delta = 0$:

$$\begin{aligned}\mathcal{C}_M(t,s) &= \mathbb{E}[M(t)\cdot M(s)] = \mathbb{E}\big\{[M(t)-M(s)+M(s)]\cdot M(s)\big\} = \\ &= \mathbb{E}\big\{[M(t)-M(s)]\cdot M(s)\big\} + \mathbb{E}\left[M(s)^2\right] = \\ &= \mathbb{E}\big\{[M(t)-M(s)]\cdot [M(s)-M(0)]\big\} + \mathbb{V}[M(s)] = \\ &= \mathbb{V}[M(s)]\,.\end{aligned}$$

Excluding the Brownian case $\alpha = 2$ (in this case we have a SBM with $\mathcal{C}_M(t,s) = \min\{t,s\}$), since the restriction $1 < \alpha < 2$ have to be respected, it leads to a L-stable random variable $M(s)$ with infinite variance. As a consequence, the covariance is also infinite

$$\mathcal{C}_M(t,s) = \infty\,.$$

Note that this process does not own long-range dependence since the mere process $\{M(t)\}_{t\in[0,+\infty)}$ is not stationary. If we consider its incremental version $\{M(t)-M(s)\}_{t\geq s\in[0,+\infty)}$, it is stationary, but its covariance is null (see Definition 4.9).

1.7 A brief summary. Noah and Joseph effects

In this chapter we have introduced three useful models to describe fluctuations of (stock) prices: Brownian motion (BM), fractional Brownian motion (FBM) and Lévy-stable motion (LSM). To be more precise, we will use them to describe the *logarithmic differences* between prices (that is a measure of rate of return). Setting $S(t)$ the stock (random) price at time t, the logarithmic return between $t-\Delta t$ and t with $0 \leq t-\Delta t < t$ is given by:

$$X(t) := \ln[S(t)] - \ln[S(t-\Delta t)]\ .$$

Each of them becomes more appropriate depending on the behaviour of single asset at hand: We want to stress that we are not looking for a *general*, *always-valid* law able to describe price movements. As we shall see, there are several stock prices which behaves *mildly* whereas other which exhibit very *wild* features. At the same time we repel the widespread doctrine which looks at the Gaussian law (and hence BM) as the only possible way to model and describe those variables. The normal measure is a very useful tool since it has many sought properties and makes easier the explicit achievement

of a lot of formulas, especially regarding derivatives pricing. On the other hand, however, the computational tools available nowadays allow us both to construct and manipulate densities that do not own a closed form and to use approximation and quadrature techniques for numerical evaluation of almost any sort of calculation. We will sum up the main properties of the three models and see in which cases they can be a useful tool for describing the variation of *certain* prices. The following dissertation is mainly based on Mandelbrot's mindset. For a more in-depth approach see [27] and [28] (but also [25] and [26]).

The economic motivation for a BM structure was given for the first time by Bachelier (1900), and at a later stage by Osborne (1959), using arguments based on the Central Limit Theorem to support the assumption of normality. If the price changes from transaction to transaction are independent, identically distributed random variables with finite variance, and if transactions are fairly uniformly spaced though time, the Central Limit Theorem leads us to believe that price changes across different intervals such as a day, a week or a month will be normally distributed since they are simple sums of the changes form transaction to transaction. Nevertheless Bachelier himself noted that this model contradicts the evidence in several ways:

(a) First, the sample variance of logarithmic return varies in time since it is observed a cyclic (non periodic) behaviour of them. Bachelier and many followers of classical finance attributed this to variability of the population variance, interpreting the sample histograms as being relative to mixtures of distinct populations;

(b) Second, he observed that the tails of the histogram could be expected to be fatter (leptokurtic) than in the Gaussian case and the distribution be asymmetric, whereas normality requires symmetry. These phenomenons materialize in discontinuous and concentrated changes. The motivation of this anomaly is often ascribed to the presence of outliers which should be ignored;

(c) Finally, stock market can exhibit short and long memory. Price changes are not always independent like in a coin tossing; it is plain for all to see that after a downfall it is more likely to register an additional drop rather than an increase. It gives rise to concentration and volatility clustering. As a matter of fact, adopting a masses' behavioural approach, we have to remember that behind records of prices there are *investors* who can be influenced by past (even far back in time) experiences and whose decision are not always rational. However, this fact is not even taken into account by many.

The Brownian motion does not face any of these concerns. On the other hand, we have seen that, in large part, the first two can be addressed by

introducing a LSM while the last one is contemplated by FBM only. Regarding the first point, if we assume that non stationarity of variance ought to be addressed to a divergent variance of the rates of returns (hence the variance is infinite), the sample variance is probably a meaningless measure of dispersion. To make matters worse, if the variance is infinite both other statistical tools (such as least-square regression) and financial equilibrium model (CAMP and APT, only to cite the most famous ones), which are based on the assumption of finite variance will be, at best, considerably weakened and may in fact give very misleading answers.

About letter **(b)**, every one must ascertain that large variations are more frequently than the ones expected if the sample were extracted by a normal population; conversely mild variations are too few. Moreover, the overall log-price changes seem to be *invariant* under different time scales: Taking into account monthly, weekly or daily diagrams, the ratio between the highest and the lowest price movements follows a regular scheme. We have seen that all three models are self-affine processes and hence able to carry out this task. Furthermore, price changes are even gathered by size: Large variations are used to come in a very rapid sequence (like a "spray of gunfire", to use a Mandelbrot's vivid expression), which are followed by smaller ones. Even here it can be traced an invariance behaviour: if we analyze at a higher frequency the group of large variations, we shall see they are composed even of small ones. Moreover at a bigger scale, we are able to find even smaller clusterings. This is a *fractal structure* that will be analyzed in the next chapter. As a consequence of self-affinity, we will prove that all the three models are (random) fractal models.

Indeed, the assumption of statistical independence of successive increments[33] is undoubtedly a simplification of reality. Incidentally, independence implies that no investor can use his knowledge of past data to increase his expected profit. But the converse, despite the thought of many, is not true: There exist processes in which the expected profit vanishes, but dependence is extremely long range.

The following table summarize the main properties of the three models (FBM is referred as having Hurst exponent $H \neq \frac{1}{2}$ and LSM with $\alpha \neq 2$).

The last part of the section is devoted to the Noah and Joseph effects, which are related to the *misbehavior* of markets we have just exposed. These names were coined by Mandelbrot and are linked to the Biblical characters.

The story of Noah may describe the first wild trait of financial markets, such as the rapid price changes and the discontinuity. As it can be found in the Genesis, when Noah was six hundred years old God ordered the Flood

[33]Independence of increments is the fundamental for principle of *non-arbitrage* and Fama's *efficient-market hypotheses*. We will discuss this problem about FBM in Chapter 4.

to purify the world, for too long dominated by evilness[34].

Table 1.1: The features of the three models.

	SBM	FBM	LSM
Continuity of paths	✓	✓	×
Differentiability of paths	×	×	×
Distribution of increments	Gaussian	Gaussian	L-stable
Independence of increments	✓	×	✓
Stationarity of increments	✓	✓	✓
Mean of the process	0	0	∞ or 0^a
Covariance of the process	finite	finite	infinite
Semi-martingale	✓	×	✓
Self-affinity	✓	✓	$\checkmark^b$
Long-range dependence	×	$\checkmark^c$	×
Heaviness of the tails	×	×	✓

[a] The mean of the process is infinite if $0 < \alpha \leq 1$ and zero if $1 < \alpha < 2$.

[b] The process is self-affine if and only if $1 < \alpha < 2$.

[c] The process has long-range dependence if and only if $\frac{1}{2} < H < 1$.

The flood-tide arrived and then went away, catastrophic but temporary. Market downfalls are very similar: The unconscionable drop of the 29.2% registered by the Dow Jones on the *Black Monday* (October 19$^{\text{th}}$, 1987) arrived unexpectedly. At that time it seemed to be at the end of the (financial) world. Mild storms hits more frequently with limited effects. Financial

[34] *"Seven days from now I will send rain on the earth for forty days and forty nights, and I will wipe from the face of the earth every living creature I have made."* Genesis, 7, 4 *"... on that day all the springs of the great deep burst forth, and the floodgates of the heavens were opened."* Genesis, 7, 11

movements are ruled by a wide hierarchy of turbulence which repeats itself at different time-scales. This is what Mandelbrot named *Noah effect*.

The second wild aspect involved in market behaviour is the presence of aperiodic pseudo-cycles, which Mandelbrot associates with the story of Joseph. The pharaoh had dreamt that, while seven sleek and fat cows were grazing, other seven ugly and gaunt cows emerged from the Nile and devoured the former. Joseph, a sensory-gifted Hebrew slave, was asked by the pharaoh to interpret the dream: Seven years of prosperity would be followed by seven years of famine. Years later, when what he had prophesied occurred, the pharaoh was able to feed his whole population having saved a lot of wheat during the elapsed time.[35] Considering all the gains Joseph and the pharaoh made, they could be addressed as the first international arbitrageurs. That peculiarity of Nile, well-known by the hydrologist Hurst, has a counterpart in financial markets which is called *Joseph effect* for obvious reasons. A 3% variation of a stock price may precede an increase of 2% the following day, then one more of 1.5% and also one of 3.5%: It as if the first shocks continued to reverberate during subsequent trading days. Of course that is not a regular or predictable structure and we are supposing it hides a long-range dependence of a stochastic process which hance owns an infinite memory.

Both effects are clear, with different intensities, looking at financial charts. Moreover we can say they are the two sides of the same coin. In order to measure the impact of those two effects, Mandelbrot elaborated several statistical tools depending on whether the underlying stochastic process is a LSM or a FBM. In the first case we will be interested in estimating the value of α (and hence the Noah effect): A market with a low value is riskier ($\alpha \to 1^+$) than one with a higher value ($\alpha \to 2^-$), being more Gaussian. If, on the other hand, we are modelling a FBM, we will fucus on the Joseph effects and thus on the H parameter: The data will be persistent, that is they tends to move in the same direction, if $H > \frac{1}{2}$ and anti-persistent if $H < \frac{1}{2}$. If $H = \frac{1}{2}$ the increments will be independent.

In order to distinguish the two effects, Mandelbrot introduced a non-parametric test called *Rescaled Range Analysis* or, more briefly, *R/S Analysis*. One of its strong point is to avoid any hypothesis regarding the probabilistic structure of data. The R/S formula measure if, in time intervals with diverse width, the difference between the maximum and minimum values of the data

[35] *"It is just as I said to Pharaoh: God has shown Pharaoh what he is about to do. Seven years of great abundance are coming throughout the land of Egypt, but seven years of famine will follow them. Then all the abundance in Egypt will be forgotten, and the famine will ravage the land."* Genesis, 41, 28-30
"When the famine had spread over the whole country, Joseph opened all the storehouses and sold grain to the Egyptians, for the famine was severe throughout Egypt. And all the world came to Egypt to buy grain from Joseph, because the famine was severe everywhere." Genesis, 41, 56-57

is higher or lower than the one expected we were facing an independent sample. A recurring non-independent sequence of gains or losses drives extreme values farther than ones realizable only by chance. Let $x_1, x_2, ..., x_n$ be the values of n successive daily (weekly/monthly/...) rates of return[36]

$$x_t := \ln(s_t) - \ln(s_{t-\Delta t})\,^{37}\,.$$

We start computing the return for the different intervals until the entire time series of n days. Then, for every interval, it is computed the difference between the corresponding rate and the sample mean. Then the cumulative values are taken, that is

$$\sum_{i=1}^{k}\left(x_i - \frac{1}{n}\sum_{i=1}^{n} x_i\right) = \sum_{i=1}^{k} x_i - \frac{k}{n}\sum_{i=1}^{n} x_i$$

with $k = 1, 2, ..., n$. Considered the whole sample, we will have a maximum and a minimum difference. Subtracting these two values we have an estimate of the *sample range* from the highest to the least peak of the deviation of the cumulative return, that is

$$\mathcal{R}_k := \max_{1\le k\le n}\left(\sum_{i=1}^{k} x_i - \frac{k}{n}\sum_{i=1}^{n} x_i\right) - \min_{1\le k\le n}\left(\sum_{i=1}^{k} x_i - \frac{k}{n}\sum_{i=1}^{n} x_i\right)\,. \tag{1.48}$$

Furthermore the empirical *standard deviation* is

$$\mathcal{S}_k := \sqrt{\frac{1}{k}\sum_{i=1}^{k}\left(x_i - \frac{1}{k}\sum_{i=1}^{k} x_i\right)^2}\,. \tag{1.49}$$

Based on the records of observations of Nile flows in 622-1469, Hurst discovered that as $n \to +\infty$

$$\frac{\mathcal{R}_k}{\mathcal{S}_k} \sim C \cdot k^H$$

for a positive constant C and $0 < H < 1$, where the latter is the usual Hurst exponent. Setting $a_k = \lim_{k\to\infty} \frac{\mathcal{R}_k}{\mathcal{S}_k}$, the estimate of H can be obtained as the slope of the log-log plot. As a matter of fact

$$\ln(a_k) = \ln(C) + H \cdot \ln(k)\,. \tag{1.50}$$

If the time series is generated by a Brownian motion process it has the value of $H = \frac{1}{2}$. If the sample range is wider than the normal case ($H > \frac{1}{2}$), data are persistent; conversely, if it is more limited ($H < \frac{1}{2}$), data are anti-persistent and always tend to change direction.

[36] We use lower cases letters instead of capital ones since we are dealing with (discrete) empirical data extracted from the population.

[37] If the data have daily frequency, we have $\Delta t = 1$ day.

Nevertheless, at first sight, we should conclude that the Hurst exponent is related to the Joseph effect only. To explain the connection with the Noah one, we will use the vivid example given by Mandelbrot in [28]. We have seen how the Joseph effect depends on the time order of the events, whereas the Noah effect is related only to their dimension. If we mix the data up, as we would do for shuffling cards, we eliminate the sequential structure and hence the Joseph effect disappears. Only the scores of the cards (the Noah effect) remain visible before and after the reshuffle. If we notice a discrepancy between the pre-shuffle and the post-shuffle, it must be caused by the long range dependence (the Joseph effect); in this case the precise time sequence is crucial. If no dissimilarity appears, thus the dependence level – if ever existed – can be considered negligible. As a consequence we succeed in obtaining a long-rang dependence measure.

Moreover, it happens that sometimes these two effect are totally linked, being $H = \frac{1}{\alpha}$. In this case (that is the LSM case with $1 < \alpha \leq 2$), from a mathematical point of view, the connection between the two effects is so deep to be a *duality*. Every random variables of a LSM gets heavy tails, being linked to a "counterpart" with long-range dependence and vice versa.

Chapter 2

Principles of fractal geometry

> *Geometry* is concerned with making our spatial intuitions objective. Classical geometry provides a first approximation to the structure of physical objects [...]. Fractal geometry is an extension of classical geometry [...]. Fractal geometry is a new language.
>
> Barnsley, M.

The inventor of fractal geometry is doubtless the French-Polish mathematician Benoît Mandelbrot who coined the word *fractal* in his book *Les Objects Fractals: Forme, Hazard et Dimension* (1975), from the Latin *fractus*, meaning "broken", to describe objects that were too irregular to fit into a traditional geometrical setting. Before him several mathematicians, such as George Cantor, Henri Poincaré and Paul Lévy mentioning only a few, hit on fractal structures but did not provide a general and theoretical treatise. This preliminary exposition aims at giving an informal introduction to fractal objects, mostly based on the show of three basic examples taken from [13] and [14].

We begin with a brief look at three simple fractals and note some of their properties. The first one is the *middle third Cantor set*, which is one of the best known and easily constructed fractal. It is constructed from a unite interval by a sequence of deletion operations. Let C_0 be the interval $[0,1]$. Let C_1 be the set obtained by deleting the middle third of C_0, so that C_1 is composed by the intervals $\left[0,\frac{1}{3}\right]$ and $\left[\frac{2}{3},1\right]$. Deleting the middle thirds of these intervals we get C_2, which is the union of $\left[0,\frac{1}{9}\right]$, $\left[\frac{2}{9},\frac{1}{3}\right]$, $\left[\frac{2}{3},\frac{7}{9}\right]$ and $\left[\frac{8}{9},1\right]$. Continuing with this procedure, C_n is given by the deletion of the middle third of each interval of C_{n-1}. Thus C_n consists of 2^n intervals each of length equal to $\frac{1}{3^n}$. The middle third Cantor set C can be defined as

$$C := \bigcap_{n=0}^{\infty} C_n \,.$$

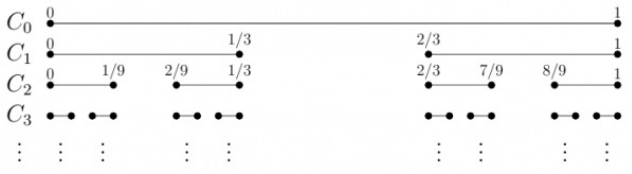


Figure 2.1: The middle third Cantor set

It may be thought as the limit of intersections of the sequence sets C_n as n tends to infinity.

This set is infinite and uncountable, containing infinitely many numbers in every neighbourhood of each of its points. Since we have seen the set C_n consists of 2^n disjoint closed intervals, each of length $\frac{1}{3^n}$, so the total length of it is equal to

$$2^n \cdot \frac{1}{3^n} = \left(\frac{2}{3}\right)^n .$$

Because of it, the total length of the middle third Cantor set can be thought as given by the limit:

$$\lim_{n\to\infty} \left(\frac{2}{3}\right)^n = 0 .$$

Hence the "classical" length is not a very useful way to compute the size of C. We will see that this is related to the fact that its fractal dimension is less then one.

Another simple example of fractal is represented by the *von Koch curve*. Let E_0 be a straight line with unit length. The set E_1 consists of the four segments obtained by removing the middle third of E_0 and replacing by other the other two sides of the equilateral triangle whose base is given by the removed segment. To construct E_2 we apply the same procedure to each segments in E_1 and so on. Thus E_n is obtained by replacing the middle third of each straight segment of E_{n-1} by the other two sides of an equilateral triangle. When n is large, the curves E_{n-1} and E_n differ only in fine detail. If n tends to infinity, the sequence of polygonal curves E_n approaches to a limiting curve E, which is the von Koch curve.

$$E := \lim_{n\to\infty} E_n .$$

This curve has features very similar to the ones seen for the middle third Cantor set. As a matter of fact E_1 is made up of four quarters each similar to the whole, but scaled by a factor $\frac{1}{3}$. Its structure is reflected in the irregularities at all scales; nevertheless, this intricate curve arises from a very simple construction. Whilst it is reasonable to call E a curve, it is much too irregular to have tangents in the classical sense. Moreover, since every straight line gets divided into four quarters having a length reduced

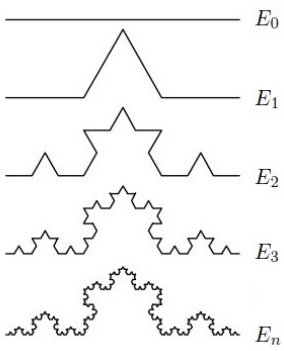


Figure 2.2: The von Koch curve

by $\frac{1}{3}$ with respect to the original size, we can heuristically state that E_n is of length equal to

$$4^n \cdot \frac{1}{3^n} = \left(\frac{4}{3}\right)^n .$$

Thus the total length of the von Koch curve, unlike Cantor set, is infinite. As a matter of fact, the exponentiation with base grater than one positively diverges, that is

$$\lim_{n\to\infty} \left(\frac{4}{3}\right)^n = \infty .$$

At the same time however, E occupies zero area in the plan, so neither length nor area provide a useful characterization of it.

The last set we show in this naive introduction is the *Sierpiński triangle*. It is obtained by repeatedly removing equilateral triangles from an initial equilateral triangle of unite side-length. Let we start with a filled-in equilateral triangle (together with the region inside), called S_0. It may be subdivided into four smaller triangles, using lines which connect the middle points of all sides, so having a side-length equal to $\frac{1}{2}$. Then the middle triangle, which is rotated by π radians compared to the others, has to be removed. After its removal, the reaming set is S_1, which is a subset of S_0. Now each of the remaining triangles should be subdivided into smaller triangles with side-length equal to $\frac{1}{4}$, and the three middle triangles removed. The result is S_2, a subset of S_1. Continuing in the same way, we get a sequence S_n of sets. The Sierpiński triangle is the limit S of this sequence of sets. The sequence is decreasing $S_0 \supset S_1 \supset ... \supset S_n \supset ...$, so by the limit we mean the intersection

$$S := \bigcap_{n=0}^{\infty} S_n .$$

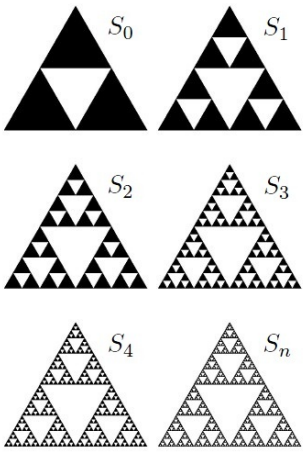


Figure 2.3: The Sierpiński triangle

The set S_n consists of 3^n triangles with side $\frac{1}{2^n}$. So the total area or S_n is

$$3^n \cdot \frac{\frac{1}{2^n} \cdot \left(\frac{1}{2^n} \cdot \frac{\sqrt{3}}{2}\right)}{2} = \frac{\sqrt{3}}{4} \cdot 3^n \cdot \left(\frac{1}{2^n}\right)^2 = \frac{\sqrt{3}}{4} \cdot \left(\frac{3}{4}\right)^n .$$

Since

$$\lim_{n\to\infty} \frac{\sqrt{3}}{4} \cdot \left(\frac{3}{4}\right)^n = 0 ,$$

the total area of the Sierpiński triangle is null. Thus the area is not a very useful tool in measuring the size of S, although *area* is usually used to measure the size of a set of two-dimension, like the Sierpiński triangle. A line segment, which is a one-dimensional object, has area equal to zero; in a similar way, we will see that the Sierpiński triangle S can be said to have dimension less than two.

Furthermore, since the set S_n contains 3^n triangles, each having three sides of length $\frac{1}{2^n}$, its total length is

$$3^n \cdot 3 \cdot \frac{1}{2^n} = 3 \cdot \left(\frac{3}{2}\right)^n .$$

Hence the total length of the Sierpiński triangle S is given by the limit of n approaching to infinity, that is

$$\lim_{n\to\infty} 3 \cdot \left(\frac{3}{2}\right)^n = \infty ,$$

since the base is grater than one. So nighter area nor length are good ways to measure the size of S. Since *length* is used to measure the size of a set of one-dimension, a filled-in square, which is a two-dimensional object, has

infinite length, because it contains infinitely many disjoint segments. In a similar way, we shall see that the Sierpiński triangle S has a dimension grater than one. Hence we will conclude its dimension is a number between one and two (hence it is more than a segment, but less than a two-dimensional filled-in figure).

In his original essay, Mandelbrot defined a fractal to be a set with Hausdorff dimension strictly grater than its topological dimension (See Definition 2.2.5). However, this definition revealed itself to be unsatisfactory since it excludes a number of sets that clearly ought to be claimed as fractals. About the puzzling definition of fractal, Falconer ([14]) provides us with a vivid simile: Asking a mathematician what a "fractal" is appears as the same thing of asking a biologist what "life" is. There is no hard and fast definition, but just a list of properties characteristic of a living being. Because of this, it seems best to regard a fractal as a set that has properties such as those listed below, rather than to look for a precise definition which will almost surely exclude other interesting cases.

When we refer to a set F as a fractal, we will typically observe (almost all of these feature will be in-depth analyzed in the following sections):

- F has a *fine* structure, that is it owns details on arbitrarily small scale;
- F is too *irregular* to be described in traditional geometrical language, both locally and globally;
- Often F has some form of *self-similarity*, not only exact but also approximate or statistical;
- Usually, its *fractal dimension* (defined in some way) is grater than its topological dimension;
- In most cases, a fractal is defined in a very simple way, usually *recursively*.

We conclude this preliminary hint to fractals remarking that circles or parabolas occur as the solution curves of certain differential equations and a knowledge of their geometrical properties help us in understanding the differential equations themselves. In the same way, the general theory of fractal geometry can be applied to many branches of mathematics and applied science in which fractals occur.

2.1 Measures of fractal sets

In this section we are interested in giving a brief discussion of the Measure Theory. We will show that the classical Lebesgue measure, in spite of being a very "refined" measure, it is not able to give a measure of fractals. Hence

we will be required to introduce another measure able to get the typical irregularities of fractals. Subsequently, that new notion of measure will be used to give a more sophisticated definition of dimension. As a matter of fact, as the Lebesgue measure is not able to measure a fractal, the definition of dimension given above is unsatisfactory for fractals as well.

Below we will focus on measures on subsets of $\mathbb{R}^n$. Basically a measure is just a way of ascribing a numerical "size" to sets, such that, if a set is decomposed into a countable number of non-overlapping subset, then the size of the whole is the sum of the sizes of the pieces. In this sense, a measure is a generalization of the concepts of length, area and volume.

Definition 2.1.1 *Let X be a set and Σ be a σ-algebra over X. A function $\mathcal{M} : \Sigma \to \mathbb{R}_0^+ \cup \{+\infty\}$ is called* measure *on X if it satisfies the following properties:*

(a) $\mathcal{M}(\varnothing) = 0$;

(b) *If $\{A_i\}_{i\in\mathbb{N}} \in \Sigma$ is a disjoint sequence of sets, then*

$$\mathcal{M}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathcal{M}(A_i)\,.$$

We call property **(b)** *countable additivity*. The triple $(X, \Sigma, \mathcal{M})$ is called a *measure space*, where X is any set, Σ is a σ-algebra and $\mathcal{M}$ is a measure over Σ. All the sets $A \in \Sigma$ are named *measurable sets*. Moreover, if $A \in \Sigma$ and $\mathcal{M}(A) = 0$, thus A is said to be a *$\mathcal{M}$–null set* (or negligible set). Eventually, the measure $\mathcal{M}$ over X is *finite* if $\mathcal{M}(X) < \infty$ (of course, if a set is bounded, it must have finite measure).

Moreover, all the measures are monotone and countably subadditive.

Theorem 2.1.2 *Let $(X, \Sigma, \mathcal{M})$ be a measure space. Thus:*

1. *Given $A, B \in \Sigma$, if $A \subseteq B$*

$$\mathcal{M}(A) \leq \mathcal{M}(B).$$

Proof: Since $B = A \cup (B \setminus A)$ and using **(b)** of Definition 2.1.1, it follows

$$\mathcal{M}(B) = \mathcal{M}(A) + \mathcal{M}(B \setminus A) \geq \mathcal{M}(A).$$

□

2. *Given $\{A_i\}_{i\in\mathbb{N}} \in \Sigma$,*

$$\mathcal{M}\left(\bigcup_{i=1}^{\infty} A_i\right) \leq \sum_{i=1}^{\infty} \mathcal{M}(A_i)\,.$$

Proof: Since

$$\bigcup_{i=1}^{\infty} A_i = A_1 \cup (A_2 \setminus A_1) \cup \big(A_3 \setminus (A_2 \cup A_1)\big) \cup ... \cup \left(A_n \setminus \bigcup_{i=1}^{n-1} A_i\right) \cup ...$$

and hence

$$\mathcal{M}\left(\bigcup_{i=1}^{\infty} A_i\right) = \mathcal{M}(A_1) + \mathcal{M}(A_2 \setminus A_1) + \mathcal{M}\big(A_3 \setminus (A_2 \cup A_1)\big) + \\ + ... + \mathcal{M}\left(A_n \setminus \bigcup_{i=1}^{n-1} A_i\right) + ...$$

Since $\mathcal{M}\left(A_j \setminus \bigcup_{i=1}^{j-1} A_i\right) \leq \mathcal{M}(A_j)$ for all j, the thesis follows.

□

Since fractals are usually defined recursively, the following theorem is crucial for making us able to compute their measure.

Theorem 2.1.3 *Let $A_1, A_2, \dots$ be a collection, even infinite, of subset of Σ. Hence:*

$$\mathcal{M}\left(\bigcup_{i=1}^{\infty} A_i\right) = \lim_{i \to \infty} \mathcal{M}(A_i) \tag{2.1}$$

if $A_1 \subseteq A_2 \subseteq ...$, and

$$\mathcal{M}\left(\bigcap_{i=1}^{\infty} A_i\right) = \lim_{i \to \infty} \mathcal{M}(A_i) \tag{2.2}$$

if $A_1 \supseteq A_2 \supseteq ...$ and $\mathcal{M}(A_1) < \infty$ (hence A_i is bounded $\forall i = 1, 2, \dots$).

Proof: Since $A_1 \subseteq A_2 \subseteq ...$, it ensues

$$A_i = A_{i-1} \cup (A_i \setminus A_{i-1})$$

leading to $\mathcal{M}(A_i) = \mathcal{M}(A_{i-1}) + \mathcal{M}(A_i \setminus A_{i-1})$ and hence $\mathcal{M}(A_i \setminus A_{i-1}) = \mathcal{M}(A_i) - \mathcal{M}(A_{i-1})$. Moreover, because

$$\bigcup_{i=1}^{\infty} A_i = A_1 \cup (A_2 \setminus A_1) \cup ... \cup (A_i \setminus A_{i-1}) \cup ...$$

it follows

$$\mathcal{M}\left(\bigcup_{i=1}^{\infty} A_i\right) = \mathcal{M}(A_1) + \big[\mathcal{M}(A_2) - \mathcal{M}(A_1)\big] + ... + \big[\mathcal{M}(A_i) - \mathcal{M}(A_{i-1})\big] + ... = \\ = \lim_{i \to \infty} \mathcal{M}(A_i)$$

which is the first point of the thesis. In addiction, if $A_1 \supseteq A_2 \supseteq ...$, we have

$$A_1 = \left(\bigcap_{i=1}^{\infty} A_i \right) \cup (A_1 \setminus A_2) \cup ... \cup (A_i \setminus A_{i+1}) \cup ...$$

and hence

$$\mathcal{M}(A_1) = \mathcal{M}\left(\bigcap_{i=1}^{\infty} A_i \right) + \left[\mathcal{M}(A_1) - \mathcal{M}(A_2) \right] + ... + \left[\mathcal{M}(A_i) - \mathcal{M}(A_{i+1}) \right] + ... =$$
$$= \mathcal{M}\left(\bigcap_{i=1}^{\infty} A_i \right) + \mathcal{M}(A_1) - \lim_{i \to \infty} \mathcal{M}(A_i)$$

from which, since $\mathcal{M}(A_1) < \infty$, it ensues

$$\mathcal{M}\left(\bigcap_{i=1}^{\infty} A_i \right) = \lim_{i \to \infty} \mathcal{M}(A_i)$$

which is the second point of the thesis.

□

At this point, we are able to introduce the *Lebesgue measure*.

Let Q be a n-dimensional interval of $\mathbb{R}^n$. It is given by the Cartesian product of n intervals of $\mathbb{R}$, that is

$$Q = [a_1, b_1] \times [a_2, b_2] \times ... \times [a_n, b_n]$$

with $a_j \leq b_j \in \mathbb{R}^*$, for all $j = 1, 2, \ldots, n$. Moreover, the n-dimensional volume of Q is defined as

$$\mathrm{vol}^n(Q) := \prod_{j=1}^{n} (b_j - a_j) \ . \tag{2.3}$$

Of course $\mathrm{vol}^1(\cdot)$ is the length $(b-a)$ of the interval $[a, b] \in \mathbb{R}$, $\mathrm{vol}^2(\cdot)$ is the area $(b_1 - a_1) \cdot (b_2 - a_2)$ of a rectangle of $\mathbb{R}^2$ and $\mathrm{vol}^3(\cdot)$ is the usual volume $(b_1 - a_1) \cdot (b_2 - a_2) \cdot (b_3 - a_3)$ of a parallelepiped of $\mathbb{R}^3$. In addiction, if there exist a a_j or b_j infinite, hence the n-dimensional volume is also infinite.

Definition 2.1.4 *Let $A \subseteq Q$, where Q is a n-dimensional interval of $\mathbb{R}^n$. Then the n-dimensional* Lebesgue measure[1] *of A is given by*[2]

$$\mathcal{L}^n(A) := \inf \left\{ \sum_{i=1}^{\infty} \mathrm{vol}^n(A_i) : A \subset \bigcup_{i=1}^{\infty} A_i \right\} \tag{2.4}$$

[1] To be precise, this is the *outer Lebesgue measure*. However, if the set is *measurable*, its outer measure coincides with its measure.

[2] Note that $\inf(\varnothing) = +\infty$ and $\sup(\varnothing) = -\infty$. As a matter of fact, according to the Completeness Axiom, being $\mathbb{R}^n$ neither bounded above (that is $\sup(\mathbb{R}^n) = +\infty$) nor bounded below (that is $\inf(\mathbb{R}^n) = -\infty$), since $\varnothing^c = \mathbb{R}^n$, the previous statement follows trivially.

where the infimum is taken over all coverings of A by n-dimensional intervals $A_i \subseteq Q, \forall i = 1, 2, \ldots$.

This is a "classical" measure that works for the all the sets of ordinary geometry. However, we shall see that it fails when we try to give a measure to a fractal set. As a matter of fact, the Lebesgue measure is unable to distinguish between a (non-empty) fractal set and the empty set.

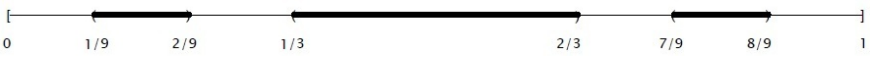
Let we work on the middle third Cantor set. It has been briefly taken into account in the introduction of this chapter since it is a very simple fractal set. In the following we will discuss about some of its properties and compute it Lebesgue measure.

To define the Cantor set, we take the interval $C_0 = [0, 1] \subset \mathbb{R}$ and we remove the *middle third* of it, that is the open interval $\left(\frac{1}{3}, \frac{2}{3}\right)$. The parts that are removed are called *tremas* (Mandelbrot coined this word from the Greek $\tau\rho\eta\mu\alpha$, meaning "hole"). After the removal, we remain with the union of

two closed intervals (note that the middle third Cantor set is represented by the non-bold intervals). Hence we have

$$C_1 = \left[0, \frac{1}{3}\right] \cup \left[\frac{2}{3}, 1\right] .$$

We go on eliminating the middle thirds of these two intervals. At this point,



we have four small closed intervals, that is

$$\begin{aligned} C_2 &= \left[0, \frac{1}{9}\right] \cup \left[\frac{2}{9}, \frac{1}{3}\right] \cup \left[\frac{2}{3}, \frac{7}{9}\right] \cup \left[\frac{8}{9}, 1\right] = \\ &= \left[0, \frac{1}{3^2}\right] \cup \left[\frac{2}{3^2}, \frac{3}{3^2}\right] \cup \left[\frac{6}{3^2}, \frac{7}{3^2}\right] \cup \left[\frac{8}{3^2}, 1\right] . \end{aligned}$$

At the following step, we get eight closed intervals even smaller, and the set becomes always more restricted.

These are

$$\begin{aligned} C_3 &= \left[0, \frac{1}{27}\right] \cup \left[\frac{2}{27}, \frac{1}{9}\right] \cup \left[\frac{2}{9}, \frac{7}{27}\right] \cup \left[\frac{8}{27}, \frac{1}{3}\right] \cup \left[\frac{2}{3}, \frac{19}{27}\right] \cup \left[\frac{20}{27}, \frac{21}{27}\right] \cup \left[\frac{8}{9}, \frac{25}{27}\right] \cup \left[\frac{26}{27}, 1\right] = \\ &= \left[0, \frac{1}{3^3}\right] \cup \left[\frac{2}{3^3}, \frac{3}{3^3}\right] \cup \left[\frac{6}{3^3}, \frac{7}{3^3}\right] \cup \left[\frac{8}{3^3}, \frac{9}{3^3}\right] \cup \left[\frac{18}{3^3}, \frac{19}{3^3}\right] \cup \left[\frac{20}{3^3}, \frac{21}{3^3}\right] \cup \left[\frac{24}{3^3}, \frac{25}{3^3}\right] \cup \left[\frac{26}{3^3}, 1\right] . \end{aligned}$$

At the fourth step, we have sixteen closed intervals. Going to infinity, we get the middle third Cantor set. One might thing that the Cantor set is composed only by its endpoints (but in fact, this is not true). Since at every step the set is contained in the previous one, that is $C_1 \supset C_2 \supset ... \supset C_i \supset ...$, the Cantor set C can be defined as:

$$C := \bigcap_{i=1}^{\infty} C_i$$

With a little more effort, it can be proved that the set can also be expressed by:

$$C = [0,1] \setminus \left(\bigcup_{i=1}^{\infty} \bigcup_{k=1}^{3^{i-1}} \left(\frac{3k-2}{3^i}, \frac{3k-1}{3^i} \right) \right)$$

Then we are interested in studying the topology of this set. First of all, the Cantor set is a *bounded* subset of the real line since $C \subset [0,1]$. Moreover it is a *closed* set, since by the definition, it is given by the countable intersection of closed set (see Theorem B.1.15). Hence it is *compact*.

A second peculiarity of this set is to be a *totally disconnected* set (it contains no intervals). In other words, taken two points of the Cantor set there is always a third point that is not inside the set. As a matter of fact, taken any open interval (a,b), if we choose i such that $\frac{1}{3^i} < b-a$, after the i-th deletion of the middle thirds, we have a disjoint union of 2^i closed intervals, whose length is equal to $\frac{1}{3^i}$, in which C must be included. Hence, since the interval (a,b) is larger than any of the disjoint close intervals of C, there exists a point of (a,b) which is not in it.

Furthermore, since no interval is in C, no point of it can be an interior point (in fact, if there existed an interior point, there would be a neighbourhood of it, and hence an open interval). Hence all the points of the Cantor set are *boundary points*. So, they might be isolated or accumulation points. However the points of C are all accumulation points: First of all we note that the points of C are either the endpoints of the tremas or a point that, at each deletion, remain inside one of the 2^i intervals left. In the first case, the point is an accumulation point for the endpoints of the following eliminated intervals. In the other case, it is easy to note that every neighbourhood of such a point is included in one of the 2^i intervals left, and hence it is an accumulation point. Thus the Cantor set is a *perfect set*.

Finally it is possible to prove that the Cantor set is an uncountable set, hence owing the *cardinality of the continuum*. Summarizing all said, we are able to state:

- $\inf(C) = \min(C) = 0$ and $\sup(C) = \max(C) = 1$;
- $\overset{\circ}{C} = \varnothing$;

- $\partial C = C$;
- $\overline{C} = C$;
- $C' = C$;
- $C^c = (-\infty, 0) \cup \left(\bigcup_{i=1}^{\infty} \bigcup_{k=1}^{3^{i-1}} \left(\frac{3k-2}{3^i}, \frac{3k-1}{3^i} \right) \right) \cup (1, +\infty)$;
- $\#(C) = \aleph_1$.

At this point, we have seen that the middle third Cantor set does not owes intervals, whereas it has the same cardinality of the real line. In a certain sense, it is as to say that, from a topological point of view, C is a "very little set", but, from the one of the cardinality, it is a "very large set" (actually, like the largest one). However, we have seen that a third way to evaluate the "size" of a set is its *measure*. Let we consider the n-th removal of middle thirds. At that stage we have a set given by 2^n disjoint closed intervals, whose length is 3^{-n}. Due to this observation, we have that the Lebesgue measure of it is given by:

$$\mathcal{L}^1(C_n) = \inf \left\{ \sum_{i=1}^{\infty} (b_i - a_i) : C_n \subset \bigcup_{i=1}^{\infty} [a_i, b_i] \right\} = 2^n \cdot \frac{1}{3^n} = \left(\frac{2}{3} \right)^n .$$

By virtue of Theorem 2.1.3, and since $C := \bigcap_{i=1}^{\infty} C_i$ and it is a compact (hence finite), we have:

$$\mathcal{L}^1(C) = \lim_{n \to \infty} \mathcal{L}^1(C_n) = \lim_{n \to \infty} \left(\frac{2}{3} \right)^n = 0 .$$

Thus, from the point of view of the (Lebesgue) measurability, we get a similar result to the one given by topological (and intuitive) argumentations. Keeping on removing the middle thirds, we have the impression to have eliminated the whole. But we know the Cantor set is not made of *nothing*: The Lebesgue measure considers it as a negligible set (more precisely a $\mathcal{L}^1$-null set), which is very similar to the empty set since it has null measure by definition. Our purpose shall be to provide a *new* measure, able to distinguish between a fractal set and the empty set.

At the same time, it is possible to prove that, given E the von Koch curve the measure of its n-th step is

$$\mathcal{L}^1(E_n) = \inf \left\{ \sum_{i=1}^{\infty} (b_i - a_i) : E_n \subset \bigcup_{i=1}^{\infty} [a_i, b_i] \right\} = \left(\frac{4}{3} \right)^n .$$

and hence the Lebesgue measure of it (since it is a closed set) is

$$\mathcal{L}^1(E) = \lim_{n \to \infty} \mathcal{L}^1(E_n) = \lim_{n \to \infty} \left(\frac{4}{3} \right)^n = \infty .$$

Our last example was the the Sierpiński triangle S, for which it can make sense calculate both length and area since it is a subset of $\mathbb{R}^2$ (also the von Koch curve is a subset of $\mathbb{R}^2$ but it area, in any way defined, is obviously zero since it is made only by straight segments). The Lebesgue measure on $\mathbb{R}$ at the n-step is given by:

$$\mathcal{L}^1(S_n) = \inf\left\{\sum_{i=1}^{\infty}(b_i - a_i) : S_n \subset \bigcup_{i=1}^{\infty}[a_i, b_i]\right\} = 3 \cdot \left(\frac{3}{2}\right)^n$$

and hence

$$\mathcal{L}^1(S) = \lim_{n\to\infty} \mathcal{L}_1(E_n) = \lim_{n\to\infty} 3 \cdot \left(\frac{3}{2}\right)^n = \infty\,.$$

Moreover, its Lebesgue measure on $\mathbb{R}^2$ (the measures its area) at the n-step is given by:

$$\mathcal{L}^2(S_n) = \inf\left\{\sum_{i=1}^{\infty}(b_{1,i} - a_{1,i}) \cdot (b_{2,i} - a_{2,i}) : S_n \subset \bigcup_{i=1}^{\infty}[a_{1,i}, b_{1,i}] \times [a_{2,i}, b_{2,i}]\right\} =$$
$$= \frac{\sqrt{3}}{4} \cdot \left(\frac{3}{4}\right)^n.$$

So the Lebesgue measure of the Sierpiński triangle is zero since

$$\mathcal{L}^2(S) = \lim_{n\to\infty} \mathcal{L}^2(E_n) = \lim_{n\to\infty} \frac{\sqrt{3}}{4} \cdot \left(\frac{3}{4}\right)^n = 0\,.$$

As for the middle third Cantor set, the Sierpiński triangle is a $\mathcal{L}^2$-null set over $\mathbb{R}^2$.

Since all these sets are not distinguished by the Lebesgue measure from the empty set (but they are not empty!), a different definition of measure is required. The measure we introduce is due to the German mathematician Felix Hausdorff, that elaborated it a long time before the emergence of fractal geometry.

Let $Q \subseteq \mathbb{R}^n$ be non-empty. We call *diameter* of Q the greatest distance apart of pair of points $\mathbf{x}, \mathbf{y}$ of the same set Q. Thus

$$\operatorname{diam}(Q) = \sup\{\|\mathbf{x} - \mathbf{y}\| : \mathbf{x}, \mathbf{y} \in Q\}\,. \tag{2.5}$$

In $\mathbb{R}^n$ a n-ball of radius r has diameter $2r$ and a hyper-cube of side length b has diameter equal to its (hyper-)diagonal, that is $b\sqrt{n}$. Note that a set Q is bounded if it has finite $\operatorname{diam}(Q)$.

Now let $\{A_i\}_{i\in\mathbb{N}}$ be a countable collection of sets of diameter at most equal to $\delta > 0$ that cover $A \subseteq \mathbb{R}^n$. This means that

$$A \subset \bigcup_{i=1}^{\infty} A_i \;:\; 0 \leq \operatorname{diam}(A_i) \leq \delta \tag{2.6}$$

for all $i \in \mathbb{N}$. If the previous statement holds, we say that $\{A_i\}_{i\in\mathbb{N}}$ is a *δ-cover* of A.

Definition 2.1.5 *Let $A \subseteq \mathbb{R}^n$ and $s \in \mathbb{R}_0^+$. For any $\delta > 0$ we define*

$$\mathcal{H}^s_\delta(A) := \inf\left\{\sum_{i=1}^{\infty}\big[\operatorname{diam}(A_i)\big]^s : \{A_i\}_{i\in\mathbb{N}} \text{ is a } \delta\text{-cover of } A\right\}. \tag{2.7}$$

Then the s-dimensional Hausdorff measure *is the measure given by*

$$\mathcal{H}^s(A) := \lim_{\delta\to 0^+} \mathcal{H}^s_\delta(A). \tag{2.8}$$

Thus we look at all covers of A by sets of diameter at most δ and seek to minimize the sum of the s-th power of the diameters. As δ decreases, the class of permissible covers of A in (2.7) is reduced. Therefore, the infimum $\mathcal{H}^s_\delta(A)$ increases and so approaches a limit as $\delta \to 0^+$. It can be proved that this limit exists for all subset A of $\mathbb{R}^n$, though the limit can be (and usually is) 0 or $+\infty$. Moreover, it is possible to show that $\mathcal{H}^s(\cdot)$ is properly a measure since $\mathcal{H}^s(\varnothing) = 0$ and $\mathcal{H}^s\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty}\mathcal{H}^s(A_i)$, with $\{A_i\}_{i\in\mathbb{N}}$ a sequence of disjoint sets.

Thus the Hausdorff measure generalizes the common idea of length, area, volume and n-dimensional volume. It may be shown that, for subsets of $\mathbb{R}^n$, the s-dimensional Hausdorff measure is just the n-dimensional Lebesgue measure multiplied for a proper factor. More precisely, if $A \subset \mathbb{R}^n$ is a Borel set[3], then

$$\mathcal{H}^s(A) = \frac{\mathcal{L}^s(A)}{b_s} \tag{2.9}$$

where b_s is the s-dimensional volume of a hyper-ball of diameter equal to one, that is:

$$b_s = \frac{\left(\frac{\pi}{4}\right)^{\frac{s}{2}}}{\Gamma\left(\frac{s}{2}+1\right)}.$$

Note that, if $s = 0, 1$, we have that the Lebesgue and Hausdorff measures coincide. Moreover, if $s = 0$, the two measures represent the number of points in A: If it is a discrete set with n elements, then $\mathcal{H}^0(A) = \mathcal{L}^0(A) = n$; whereas, if A is infinite, we have $\mathcal{H}^0(A) = \mathcal{L}^0(A) = \infty$.

The following proposition shows that, given a set A, computing its Hausdorff measure is rather more complex than its Lebesgue measure.

[3]A *Borel set* is any set in a topological space that can be formed from open sets (or, equivalently, from closed sets) through the operations of countable union, countable intersection, and (relative) complement. For a topological space X, the collection of all Borel sets on X forms a σ-algebra, known as the *Borel σ-algebra*. The Borel σ-algebra on X is the smallest σ-algebra containing all open sets (or, equivalently, all closed sets).

Proposition 2.1.6 *Given a set $A \subseteq \mathbb{R}^n$, there exists $s^* \in \mathbb{R}_0^+$ such that:*

$$\mathcal{H}^s(A) = \begin{cases} \infty & \text{if } 0 \leq s < s^* \\ h & \text{if } s = s^* \\ 0 & \text{if } s > s^* \end{cases} \tag{2.10}$$

where h is a positive real number or ∞.

Proof: If we consider that $\mathcal{H}_\delta^s(A)$ given by (2.7) is non-increasing with s, so by (2.8), $\mathcal{H}^s(A)$ is also non-increasing. Moreover, since $\{A_i\}_{i\in\mathbb{N}}$ is a δ-cover of A we have

$$\sum_{i=1}^{\infty}\big[\operatorname{diam}(A_i)\big]^s \leq \sum_{i=1}^{\infty}\big[\operatorname{diam}(A_i)\big]^{s-s^*}\cdot\big[\operatorname{diam}(A_i)\big]^{s^*} \leq \delta^{s-s^*}\sum_{i=1}^{\infty}\big[\operatorname{diam}(A_i)\big]^{s^*}.$$

Then, since infimum is a linear and increasing operator, taking the first and last member of the inequality, we have

$$\inf\left\{\sum_{i=1}^{\infty}\big[\operatorname{diam}(A_i)\big]^s\right\} \leq \inf\left\{\delta^{s-s^*}\sum_{i=1}^{\infty}\big[\operatorname{diam}(A_i)\big]^{s^*}\right\},$$

implying

$$\inf\left\{\sum_{i=1}^{\infty}\big[\operatorname{diam}(A_i)\big]^s\right\} \leq \delta^{s-s^*}\cdot\inf\left\{\sum_{i=1}^{\infty}\big[\operatorname{diam}(A_i)\big]^{s^*}\right\},$$

and

$$\mathcal{H}_\delta^s(A) \leq \delta^{s-s^*}\cdot\mathcal{H}_\delta^{s^*}(A).$$

Letting $\delta \to 0^+$, we see that if $\mathcal{H}^{s^*}(A) < \infty$, then $\mathcal{H}^s(A) = 0$ for $s > s^*$.

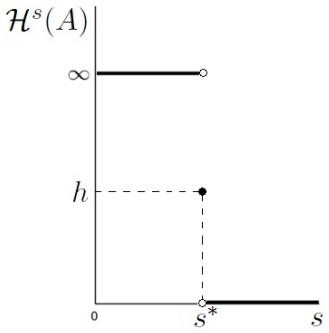


Figure 2.4: The Hausdorff dimension

Moreover, for $0 \leq s < s^*$, when δ approaches to zero, $\mathcal{H}^s(A)$ becomes infinite. Thus, considering $\mathcal{H}^s(A)$ as a function of s, it shows that there is

neighborhood of a critical value s^* at which $\mathcal{H}^s(A)$ jumps from ∞ to zero (hence in $s = s^*$ there is a *jump discontinuity*). Note that h is finite if $\mathcal{H}^{s^*}(A) < \infty$, and it is infinite if $\mathcal{H}^{s^*}(A)$ is infinite as well.

□

Since the supremum and the infimum of $\mathcal{H}^s(A)$ are equal for all sets $A \subseteq \mathbb{R}^n$, we are interested in the value of s^* (and hence the corresponding measure h), since it is different from set to set. However, to be able to understand properly what those quantities stand for, we need a more general concept of *dimension*, which will be introduced in the following section.

2.2 Dimensions of fractal sets

The notion of *dimension* has been mentioned in the first Section for vector spaces, saying that the dimension of a vector space is the cardinality of its bases. Hence we know that the Euclidean space $\mathbb{R}^n$ has $\dim(\mathbb{R}^n) = n$, since every basis is composed by n (linearly independent) vectors. In order to take into account fractal sets, it will be convenient to extend this concept of dimension to subsets of the Euclidean space.

Let $A_1, A_2, ..., A_i, ...$ be closed subsets of $\mathbb{R}^n$. We indicated the dimension of the set A_i with $\dim(A_i)$. A good definition of dimension should reasonably satisfy the following properties:

(a) $A_i \subseteq \mathbb{R}^n$ open and non-empty $\Rightarrow$ $\dim(A_i) = n$;

(b) $\#(A_i) = \aleph_0 \Rightarrow \dim(A_i) = 0$

(c) $A_i \subseteq A_j \Rightarrow \dim(A_i) \leq \dim(A_j)$ (monotonicity);

(d) $A = \bigcup_{i=1}^{\infty} A_i \Rightarrow \dim(A) = \sup_{i \geq 1} \{\dim(A_i)\}$ (σ-stability).

(e) $\dim(\cdot)$ is invariant under translations, rotations and similarities.

Hence the natural extension for subsets of $\mathbb{R}^n$ is represented by the *topological dimension.*

Definition 2.2.1 *The* topological dimension *of $A \subset \mathbb{R}^n$ is defined inductively as follows :*

- *If $A = \varnothing$, we have $\dim_{top}(\varnothing) = -1$;*
- *For an integer $k \geq 0$ and $A \neq \varnothing$, we have $\dim_{top}(A) \leq k$ if and only if there exists an arbitrarily small neighborhood $U \supseteq B(\mathbf{x}, r)$ for all $\mathbf{x} \in A$, such that $\dim_{top}(\partial U) \leq k - 1$. Moreover, we say that*

$$\dim_{top}(A) = k$$

if and only if $\dim_{top}(A) \leq k$ and $\dim_{top}(A) \not\leq k - 1$;

- *If* $\dim_{top}(A) \not\leq k$ *for all* $k \geq 0$*, then* $\dim_{top}(A) = \infty$.

In other words, the topological dimension of a set A is equal to 0, if it is totally disconnected; it is equal to 1, if for every point $\mathbf{x} \in A$ there exists an arbitrarily small neighborhood $U \supseteq B(\mathbf{x}, r)$ whose frontier ∂U has topological dimension $\dim_{top}(\partial U) = 0$; it is equal to 2, if for every point $\mathbf{x} \in A$ there exists an arbitrarily small neighborhood $U \supseteq B(\mathbf{x}, r)$ whose frontier ∂U has topological dimension $\dim_{top}(\partial U) = 1$; and so on...

We have just seen that, by definition, given $A \subset \mathbb{R}^n$, we have $\dim_{top}(A) = 0$ if and only if A is totally disconnected. Because of it, the middle third Cantor set C has zero topological dimension:

$$\dim_{top}(C) = 0\,.$$

Regarding the von Koch curve E and the Sierpiński triangle S, it obviously follows that

$$\dim_{top}(E) = 1 \quad \text{and} \quad \dim_{top}(S) = 1\,.$$

However, the topological dimension of objects is not sensitive enough of a measure to describe the intrinsic properties of fractals. In fact, using the notion of topological dimension, we essentially get a *lower bound* for the dimension of a set. First of all, note that it does not satisfy the property **(b)** of the previous list: As a matter of fact, we have seen that the Cantor set owns the cardinality of the continuum, but its topological dimension is zero. Moreover, it is neither σ-stabile, since property **(d)** fails. To show it, let we consider the set of rational numbers $\mathbb{Q}$ and its complementary $\mathbb{I}$, which is the set of irrational numbers. Of course we have $\mathbb{Q} \cup \mathbb{I} = \mathbb{R}$. By Definition B.1.18, we have that both $\mathbb{Q}$ and $\mathbb{I}$ are totally disconnected since their points can be separated by the empty set (they possess "holes"). Because of it

$$\dim_{top}(\mathbb{Q}) = \dim_{top}(\mathbb{I}) = 0$$

but

$$\dim_{top}(\mathbb{Q} \cup \mathbb{I}) = \dim_{top}(\mathbb{R}) = 1 \neq \max\{\dim_{top}(\mathbb{Q}), \dim_{top}(\mathbb{I})\}$$

since the last one is zero. Hence the topological dimension of a set does neither distinguish between the set of a discrete set and the middle third Cantor set nor possesses all the properties listed above. Thus, before introducing a dimension based on the Hausdorff measure, we define a type of *fractal dimension*[4] based on the Lebesgue measure.

Let $A \subseteq \mathbb{R}^n$ and consider a dilation on it with scaling factor $\lambda > 0$. Hence we obtain the set

$$\lambda A := \{\lambda \mathbf{x} \in \mathbb{R}^n : \mathbf{x} \in A\}\,.$$

[4]In the following exposition, we will refer to *fractal dimension* as any definition of dimension different from the topological one.

Its n-dimensional Lebesgue measure is given by[5]

$$\begin{aligned}
\mathcal{L}^n(\lambda A) &= \inf\left\{\sum_{i=1}^{\infty} \mathrm{vol}^n\left(\lambda A_i\right) : \lambda A \subset \bigcup_{i=1}^{\infty} \lambda A_i\right\} = \\
&= \inf\left\{\sum_{i=1}^{\infty} \lambda^n \cdot \mathrm{vol}^n\left(A_i\right) : A \subset \bigcup_{i=1}^{\infty} A_i\right\} = \\
&= \inf\left\{\lambda^n \cdot \sum_{i=1}^{\infty} \mathrm{vol}^n\left(A_i\right) : A \subset \bigcup_{i=1}^{\infty} A_i\right\} = \\
&= \lambda^n \cdot \inf\left\{\sum_{i=1}^{\infty} \mathrm{vol}^n\left(A_i\right) : A \subset \bigcup_{i=1}^{\infty} A_i\right\} = \\
&= \lambda^n \cdot \mathcal{L}^n(A)\,.
\end{aligned}$$

highlighting the connection existing between the dimension of the space $\mathbb{R}^n$ and the measure itself. Thus, if $\mathcal{L}^n(A) > 0$ and A is given by the union of m disjoint pairs of itself, the ones are scaled by a factor $0 < \lambda < 1$ (hence A gets reduced), by Definition 2.1.1 we have (note that $A_i = \lambda A, \forall i = 1, 2, \ldots, m$):

$$\begin{aligned}
\mathcal{L}^n(A) &= \mathcal{L}^n\left(\bigcup_{i=1}^{m} A_i\right) = \sum_{i=1}^{m} \mathcal{L}^n\left(A_i\right) = \sum_{i=1}^{m} \mathcal{L}^n\left(\lambda A\right) = \\
&= \sum_{i=1}^{m} \lambda^n \cdot \mathcal{L}^n\left(A\right) = m \cdot \lambda^n \cdot \mathcal{L}^n\left(A\right)\,.
\end{aligned}$$

Divining both members by $\mathcal{L}^n\left(A\right)$, it follows:

$$m \cdot \lambda^n = 1\,.$$

Since we are interested in the dimension n, we obtain:

$$n = -\frac{\log(m)}{\log(\lambda)}\,.$$

Note that this notion of dimension *does not depend* on the Lebesgue measure and that the base of the logarithm can be chosen in the most convenient way. Moreover, we are now able to give a consistent definition of dimension for those sets that are self-similar (see Definition 2.5.6).

[5] Since the hyper-volume of a n-dimensional interval Q is given by $\mathrm{vol}^n(Q) = \prod_{j=1}^{n}\left(b_j - a_j\right)$, a dilation on it enlarges its volume as:

$$\mathrm{vol}^n(\lambda Q) = \prod_{j=1}^{n} \lambda \cdot \left(b_j - a_j\right) = \lambda^n \cdot \prod_{j=1}^{n}\left(b_j - a_j\right)\,.$$

Definition 2.2.2 *Let $A \subseteq \mathbb{R}^n$ be the union of $m > 1$ disjoint pairs of itself scaled by a factor $0 < \lambda < 1$. We call* similarity dimension *of A the real non-negative number*

$$\dim_{sim}(A) := -\frac{\log(m)}{\log(\lambda)}\,. \tag{2.11}$$

Let we see how this new definition of dimension is really satisfactory for fractal sets. If we take into account the middle third Cantor set C, we see that, at the k–th stage, each of it segment need to be multiplied by a factor equal to $\frac{1}{3}$ to get 2 scaled copies of it at the $(k+1)$–th stage. Thus

$$\dim_{sim}(C) = -\frac{\log(2)}{\log\left(\frac{1}{3}\right)} = \frac{\log(2)}{\log(3)}\,.$$

Moreover, for the von Koch cure E we have that, at the k–th stage, every segments of the curve is reduced by $\frac{1}{3}$, leading to 4 scaled copies of it at the $(k+1)$–th stage; regarding the Sierpiński triangle S, at each stage, every sides of the triangle is scaled by $\frac{1}{2}$ obtaining 3 copies of the original one at the following stage. Then

$$\dim_{sim}(E) = \frac{\log(4)}{\log(3)} \quad \text{and} \quad \dim_{sim}(S) = \frac{\log(3)}{\log(2)}\,.$$

It is interesting to note that

$$0 < \dim_{sim}(C) < 1 \quad , \quad 1 < \dim_{sim}(E) < 2 \quad \text{and} \quad 1 < \dim_{sim}(S) < 2$$

representing the fact we have mentioned in the introduction of this chapter, asserting that a fractal set is a set whose fractal dimension (in this case, the similarity dimension) is always grater than its topological dimension (that is, respectively, zero for C and one for E and S). Furthermore, taken for instance the von Koch curve, this new notion of dimension allows as to state that such a curve is an "object" that is more than a line but less than an area, since its similarity dimension is $1.2618... \in (1,2)$.

The last effort of this section is to introduce a further fractal dimension based on the Hausdorff measure. In fact, the similarity dimension applies only to a too limited class of fractals, that are those rigorously self-similar. Many fractals found in practical applications are approximatively self-similar or statistically self-affine (see Chapter 1, Section 3). Moreover, despite its name, the similarity dimension does not satisfy the dimension conditions previously listed. However, under certain circumstances, we will see that similarity dimension coincides with an other dimension that satisfies all the required properties.

With the aid of the s-dimensional Hausdorff measure, we can now define the *Hausdorff dimension*.

Definition 2.2.3 *We define* Hausdorff dimension *of a set* $A \subset \mathbb{R}^n$ *the following quantity:*

$$\dim_{\mathcal{H}}(A) := \sup\{s : \mathcal{H}^s(A) > 0\} = \sup\{s : \mathcal{H}^s(A) = \infty\}\,. \tag{2.12}$$

By virtue of Proposition 2.1.6, this definition is the same as saying that

$$\dim_{\mathcal{H}}(A) = \inf\{s : \mathcal{H}^s(A) = 0\} = \inf\{s : \mathcal{H}^s(A) < \infty\} \tag{2.13}$$

since the function $\mathcal{H}^s(A)$ possesses a jump discontinuity. Because of it, it follows that

$$\dim_{\mathcal{H}}(A) = s^* \tag{2.14}$$

where s^* is the abscissa of the jump discontinuity of Proposition 2.1.6 (see Figure 2.4). Hence, to find the Hausdorff dimension of a set is the same thing as looking for the point at which the Hausdorff measure of the set jumps from ∞ to zero.

As expected, the Hausdorff dimension satisfies all the desirable properties a dimension should own.

Theorem 2.2.4 *Given* $A_1, A_2, ..., A_i, ...$ *subset of* $\mathbb{R}^n$*, then the following statements hold true:*

- *The Hausdorff dimension is a* monotone *function.*

  Proof: If $A_i \subseteq A_j$, from the measure property we have $\mathcal{H}^s(A_i) \leq \mathcal{H}^s(A_j)$ for each s. Hence, by the definition of Hausdorff dimension $\dim_{\mathcal{H}}(A_i) \leq \dim_{\mathcal{H}}(A_j)$.

  □

- *The Hausdorff dimension is a* σ-stable *function.*

  Proof: If $A_1, A_2, \ldots, A_i, \ldots$ is a countable sequence of sets, then certainly

  $$\dim_{\mathcal{H}}\left(\bigcup_{i=1}^{\infty} A_i\right) \geq \dim_{\mathcal{H}}(A_j)$$

  for all j from the monotonicity property. Hence, taking supremum over all j

  $$\dim_{\mathcal{H}}\left(\bigcup_{j=1}^{\infty} A_j\right) \geq \sup_{j\geq 1}\{\dim_{\mathcal{H}}(A_j)\}\,.$$

  On the other hand, if $s > \dim_{\mathcal{H}}(A_i)$ for all i, then $\mathcal{H}^s(A_i) = 0$, and so $\mathcal{H}^s(\bigcup_{i=1}^{\infty} A_i) = 0$. Hence

  $$\dim_{\mathcal{H}}\left(\bigcup_{i=1}^{\infty} A_i\right) < s$$

and then, taking the supremum over j

$$\dim_{\mathcal{H}}\left(\bigcup_{j=1}^{\infty} A_j\right) \leq \sup_{j\geq 1}\left\{\dim_{\mathcal{H}}\left(A_j\right)\right\} .$$

Because the two inequalities are both valid, it must be

$$\dim_{\mathcal{H}}\left(\bigcup_{j=1}^{\infty} A_j\right) = \sup_{j\geq 1}\left\{\dim_{\mathcal{H}}\left(A_j\right)\right\} .$$

□

- *A countable set has null Hausdorff dimension.*

  Proof: If A_i is a single point, then $\mathcal{H}^0\left(A_i\right) = 1$ and $\dim_{\mathcal{H}}\left(A_i\right) = 0$. So, by the σ-stability, $\dim_{\mathcal{H}}\left(\bigcup_{i=1}^{\infty} A_i\right) = 0$.

  □

- *If $A_i \subseteq \mathbb{R}^n$ is* open*, then its Hausdorff dimension is equal to n.*

  Proof: Since A_i contains a n-ball of positive n-dimensional volume, then $\dim_{\mathcal{H}}\left(A_i\right) \geq n$. But, since A_i is contained in countably many balls, it follows $\dim_{\mathcal{H}}\left(A_i\right) \leq n$. Thus

  $$\dim_{\mathcal{H}}\left(A_i\right) = n\,.$$

  □

- *The Hausdorff measure is invariant under similarities* (*see* Definition 2.3.2). *Furthermore, the Hausdorff dimension is invariant under bi-Lipschitz transformations* (*see* Definition 2.3.1).

  Proof: The proof of it will be given in the following section (see Proposition 2.5.4 and Corollary C.2.3).

At this point, we would be expected to compute the Hausdorff dimension of the three fractal set we have introduce previously. However, we shall not do it immediately, since the Hausdorff dimension is not so easy and immediate to compute. Fortunately, in the following section we will state crucial theorem that allow to make a connection between Hausdorff and similarity dimensions.

On the other hand, we are now able to give a general and formal definition of *fractal*.

Definition 2.2.5 *We say that a set $A \subset \mathbb{R}^n$ is* fractal *if*

$$\dim_{top}(A) < \dim_{\mathcal{H}}(A)\,.$$

2.3 Self-similarity and self-affinity

We have seen that a self-similar set is loosely speaking a set consisting of scaled copies of itself. Here, we shall give a formal definition and an in-depth analysis. We need some preliminary definitions

Definition 2.3.1 *For a closed set $A \subset \mathbb{R}^n$, the mapping $\mathbf{F} : A \to \mathbb{R}^m$ is called a* Hölder function *of exponent $\beta > 0$ on A, if there exists a constant $\lambda > 0$ such that*

$$\|\mathbf{F}(\mathbf{x}) - \mathbf{F}(\mathbf{y})\| \leq \lambda \cdot \|\mathbf{x} - \mathbf{y}\|^{\beta} \tag{2.15}$$

for all $\mathbf{x}, \mathbf{y} \in A$. If β may be taken equal to 1, $\mathbf{F}$ is called a Lipschitz function. *That is*

$$\|\mathbf{F}(\mathbf{x}) - \mathbf{F}(\mathbf{y})\| \leq \lambda \cdot \|\mathbf{x} - \mathbf{y}\| \tag{2.16}$$

for all $\mathbf{x}, \mathbf{y} \in A$. Moreover, if $\lambda < 1$, $\mathbf{F}$ is called a contraction *on A.*

If there exist two constants $0 < \lambda_1 \leq \lambda_2$, such that

$$\lambda_1 \cdot \|\mathbf{x} - \mathbf{y}\| \leq \|\mathbf{F}(\mathbf{x}) - \mathbf{F}(\mathbf{y})\| \leq \lambda_2 \cdot \|\mathbf{x} - \mathbf{y}\|$$

for all $\mathbf{x}, \mathbf{y} \in A$, the mapping is named a bi-Lipschitz function.

The smallest $(1 >) \lambda$ satisfying (2.16) is called the (*contraction*) *ratio* of $\mathbf{F}$. Moreover, a contraction is a continuous mapping.

Definition 2.3.2 *Given a closed set $A \subset \mathbb{R}^n$, we say that the mapping $\mathbf{F} : A \to \mathbb{R}^m$ is a* similarity *if* (2.16) *holds with equality, that is*

$$\|\mathbf{F}(\mathbf{x}) - \mathbf{F}(\mathbf{y})\| = \lambda \cdot \|\mathbf{x} - \mathbf{y}\| . \tag{2.17}$$

Moreover, if $\lambda < 1$, we call the mapping contracting similarity.

In other words, a similarity is an application obtained by the composition of an *homothety* (that is a transformation that enlarges/contracts an object, leaving its angles unchanged), and an *isometry* (that is the rigid movement of an object – translations, rotations and reflections).

Now we are able to introduce the concept of *invariant* and *self-similar set*.

Definition 2.3.3 *For a family $\mathfrak{F} = \{\mathbf{F_1}, \mathbf{F_2}, ..., \mathbf{F_k}\}$ of contractions with ratios $\lambda_1, \lambda_2, ..., \lambda_k$, such that $\mathbf{F_i} : A \to \mathbb{R}^m$ for all $i = 1, 2, \ldots, k$, we call a non-empty compact subset $B \subseteq A$ an* invariant set *for $\mathfrak{F}$ if*

$$B = \bigcup_{i=1}^{k} \mathbf{F_i}(B) .$$

It can be proved (see [14]) that for any $\mathfrak{F}$ there is a unique invariant set. If the elements of $\mathfrak{F}$ are similar contractions, we are able to state the following definition.

Definition 2.3.4 *For a family* $\mathfrak{F} = \{\mathbf{F_1}, \mathbf{F_2}, ..., \mathbf{F_k}\}$ *of contracting similarities defined as in* (2.17)*, the invariant set of such a collection of contracting similarities is called a* self-similar set.

Alternatively speaking, the definition implies that the set $B \subset \mathbb{R}^n$ is self-similar, if the following expression

$$\|\mathbf{F_i}(\mathbf{x}) - \mathbf{F_i}(\mathbf{y})\| = \lambda_i \cdot \|\mathbf{x} - \mathbf{y}\| . \tag{2.18}$$

holds true for all $i = 1, 2, ..., k$ and $\mathbf{x}, \mathbf{y} \in B$.

An important further condition is required to connect similarity dimension and Hausdorff dimension.

Definition 2.3.5 *We say that the contractions* $\mathbf{F_1}, \mathbf{F_2}, ..., \mathbf{F_k}$ *fulfill the* open set condition *if there exist a non-empty bounded open set* $E \subset \mathbb{R}^n$ *such that*

$$\bigcup_{i=1}^{k} \mathbf{F_i}(E) \subset E$$

with the $\mathbf{F_i}(E)$ *pairwise disjoint.*

Then the following theorem (for its proof see [14]) allows us to make a connection between the Hausdorff dimension and the similarity one.

Theorem 2.3.6 *Let* $\mathbf{F_i}(E)$ *be contracting similarities on* $\mathbb{R}^n$ *satisfying the open set condition with contraction ratios* $\lambda_i, i = 1, 2, ..., k$*. If* E *is the invariant set of* $\{\mathbf{F_i}\}_{i=1}^{k}$*, then*

- $\dim_{sim}(E) = \dim_{\mathcal{H}}(E)$*;*
- $0 < \mathcal{H}^{s^*}(E) < \infty$*, where* $s^* = \dim_{\mathcal{H}}(E)$*;*
- *There exist* $e_1, e_2 \in \mathbb{R}^+$ *such that, for* $s^* = \dim_{\mathcal{H}}(E)$*,*

$$e_1 \cdot r^{s^*} \leq \mathcal{H}^{s^*}(E \cap B(\mathbf{x}, r)) \leq e_2 \cdot r^{s^*}$$

for all $\mathbf{x} \in E$ *and* $0 < r \leq 1$.

If the open set condition is not fulfilled, then we instead get the relation $\dim_{sim}(E) \geq \dim_{\mathcal{H}}(E)$.

Since Hausdorff dimension is more difficult to compute than the similarity dimension, if we are dealing with a self-similar set, for which the open set condition holds true, we can compute its similarity dimension instead of the Hausdorff one.

Moreover, because of the Hausdorff dimension satisfies all the properties a dimension ought to have, even similarity dimension of self-similar sets fulfills

them. Hence, for the middle third Cantor set, the von Koch curve and the Sierpiński triangle we find

$$\dim_{\mathcal{H}}(C) = \frac{\log(2)}{\log(3)} \quad , \quad \dim_{\mathcal{H}}(E) = \frac{\log(4)}{\log(3)} \quad \text{and} \quad \dim_{\mathcal{H}}(S) = \frac{\log(3)}{\log(2)} .$$

Then we state and prove some results about Hausdorff measure and dimension over self-similar sets and similarities. We begin with the *scaling property*.

Proposition 2.3.7 *Let* $\mathbf{F}$ *be a similarity transformation of scale ratio* $\lambda > 0$. *If* $A \subset \mathbb{R}^n$, *then*

$$\mathcal{H}^s[\mathbf{F}(A)] = \lambda^s \cdot \mathcal{H}^s(A) .$$

Proof: If $\{A_i\}_{i\in\mathbb{N}}$ is a δ-cover of A, then $\{\mathbf{F}(A_i)\}_{i\in\mathbb{N}}$ is a $\lambda\delta$-cover of $\mathbf{F}(A)$, so

$$\sum_{i=1}^{\infty} \{\mathrm{diam}[\mathbf{F}(A)]\}^s \leq \sum_{i=1}^{\infty} \{\mathrm{diam}[\mathbf{F}(A_i)]\}^s$$

$$\sum_{i=1}^{\infty} [\mathrm{diam}(\lambda \cdot A)]^s \leq \sum_{i=1}^{\infty} [\mathrm{diam}(\lambda \cdot A_i)]^s = \lambda^s \cdot \sum_{i=1}^{\infty} [\mathrm{diam}(A_i)]^s .$$

Taking the infimum, we get

$$\mathcal{H}^s_{\lambda\delta}(\lambda \cdot A) \leq \lambda^s \cdot \mathcal{H}^s_{\delta}(A) .$$

Letting $\delta \to 0^+$, it gives that

$$\mathcal{H}^s(\lambda \cdot A) \leq \lambda^s \cdot \mathcal{H}^s(A)$$
$$\mathcal{H}^s[\mathbf{F}(A)] \leq \lambda^s \cdot \mathcal{H}^s(A) .$$

Replacing λ with $\frac{1}{\lambda}$ and A with $\lambda \cdot A$, the previous inequalities become

$$\mathcal{H}^s\left(\frac{1}{\lambda} \cdot \lambda \cdot A\right) \leq \left(\frac{1}{\lambda}\right)^s \cdot \mathcal{H}^s(\lambda \cdot A)$$
$$\lambda^s \cdot \mathcal{H}^s(A) \leq \mathcal{H}^s(\lambda \cdot A)$$
$$\lambda^s \cdot \mathcal{H}^s(A) \leq \mathcal{H}^s[\mathbf{F}(A)] .$$

Since both inequalities hold, it must be

$$\mathcal{H}^s[\mathbf{F}(A)] = \lambda^s \cdot \mathcal{H}^s(A) .$$

□

By this proposition, it follow naturally that, if $\mathbf{F}$ is an isometry (that is a similarity with $\lambda = 1$), hence

$$\mathcal{H}^s[\mathbf{F}(A)] = \mathcal{H}^s(A)\,.$$

Thus, Hausdorff measure is *translation*, *rotation* and *reflection invariant*[6], as it would certainly be expected.

On the other hand, we expect the Hausdorff dimension to be invariant. The following proposition explains this.

Proposition 2.3.8 *For a closed set $A \subset \mathbb{R}^n$, let the mapping $\mathbf{F} : A \to \mathbb{R}^m$ be a Hölder function of exponent β and ratio λ. Then, for each s*

$$\mathcal{H}^{\frac{s}{\beta}}[\mathbf{F}(A)] \leq \lambda^{\frac{s}{\beta}} \cdot \mathcal{H}^s(A)$$

and

$$\dim_{\mathcal{H}}[\mathbf{F}(A)] \leq \frac{1}{\beta} \cdot \dim_{\mathcal{H}}(A)\,.$$

Proof: If $\{A_i\}_{i\in\mathbb{N}}$ is a δ-cover of A, then, since

$$\operatorname{diam}\left[\mathbf{F}\left(A \cap A_i\right)\right] \leq \lambda \cdot \left[\operatorname{diam}\left(A \cap A_i\right)\right]^{\beta} \leq \lambda \cdot \left[\operatorname{diam}\left(A_i\right)\right]^{\beta}$$
$$\operatorname{diam}\left[\mathbf{F}\left(A \cap A_i\right)\right]^{\frac{s}{\beta}} \leq \lambda^{\frac{s}{\beta}} \cdot \left[\operatorname{diam}\left(A_i\right)\right]^{s}$$

it follows that $\{\mathbf{F}\left(A \cap A_i\right)\}_{i\in\mathbb{N}}$ is a $\lambda\delta^{\beta}$-cover of $\mathbf{F}(A)$. Thus

$$\sum_{i=1}^{\infty} \operatorname{diam}\left[\mathbf{F}\left(A \cap A_i\right)\right]^{\frac{s}{\beta}} \leq \lambda^{\frac{s}{\beta}} \sum_{i=1}^{\infty} \left[\operatorname{diam}\left(A_i\right)\right]^{s}\,.$$

Taking the infimum and setting $\delta \to 0^+$, we have the first result

$$\mathcal{H}^{\frac{s}{\beta}}[\mathbf{F}(A)] \leq \lambda^{\frac{s}{\beta}} \cdot \mathcal{H}^s(A)\,.$$

Secondly, if $s > \dim_{\mathcal{H}}(A)$, it follows $\mathcal{H}^s(A) = 0$. Because of by the definition of Hausdorff dimension we have

$$\dim_{\mathcal{H}}[\mathbf{F}(A)] = \inf\left\{\frac{s}{\beta} : \mathcal{H}^{\frac{s}{\beta}}[\mathbf{F}(A)] = 0\right\}$$

[6] As a matter of fact, the functions

$$\mathbf{F} : \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \mapsto \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} \quad \text{(translation)}$$
$$\mathbf{F} : \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \mapsto \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \quad \text{(rotation)}$$
$$\mathbf{F} : \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \mapsto \begin{pmatrix} \cos 2\theta & \sin 2\theta \\ \sin 2\theta & -\cos 2\theta \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \quad \text{(reflection)}$$

are all isometries in the plane.

implying that

$$\dim_{\mathcal{H}}[\mathbf{F}(A)] \leq \frac{1}{\beta} \cdot \dim_{\mathcal{H}}(A)$$

that is the second result.

□

Easily, the following corollary holds true.

Corollary 2.3.9 *If the mapping* $\mathbf{F} : \mathbb{R}^n \supset A \to \mathbb{R}^m$ *is a Lipschitz function, then*

$$\mathcal{H}^s[\mathbf{F}(A)] \leq \lambda^s \cdot \mathcal{H}^s(A)$$

and

$$\dim_{\mathcal{H}}[\mathbf{F}(A)] \leq \dim_{\mathcal{H}}(A)\,.$$

Moreover, if $\mathbf{F} : \mathbb{R}^n \supset A \to \mathbb{R}^m$ *is a bi-Lipschitz function, then*

$$\dim_{\mathcal{H}}[\mathbf{F}(A)] = \dim_{\mathcal{H}}(A)\,.$$

In particular, $\dim_{\mathcal{H}}[\mathbf{F}(A)] \leq \dim_{\mathcal{H}}(A)$ holds for any differentiable function $\mathbf{F}$ with bounded derivative; such a function is necessarily Lipschitz as a consequence of the mean value theorem.

Moreover, this corollary reveals a fundamental property of Hausdorff dimension: Hausdorff dimension is *invariant under bi-Lipschitz transformations*. Thus if two sets have different dimensions there cannot be a bi-Lipschitz mapping from one onto the other. In fact, one approach to fractal geometry is to regard two sets as "the same" if there is a bi-Lipschitz mapping between them.

For our proposals, a more general type of plane transformation will be necessary, since similarity applies the same scale factor in the same direction. The natural extension is given by *self-affine sets*, which form an important class of sets, including self-similar sets as a particular case.

Definition 2.3.10 *For a closed set* $A \subset \mathbb{R}^n$*, the mapping* $\mathbf{S} : A \to \mathbb{R}^m$ *is called a* affine transformation *if it can be written as*

$$\mathbf{S}(\mathbf{x}) = \mathbf{T}(\mathbf{x}) + \mathbf{b} \tag{2.19}$$

where $\mathbf{T}(\mathbf{x})$ *is a linear transformation*[7] *on* $\mathbb{R}^n$ *and* $\mathbf{b}$ *is a vector in* $\mathbb{R}^m$*.*

Thus an affine transformation is a combination of a translation, rotation, dilation and, perhaps, a reflection. Unlike similarities, affine transformations contract with different ratios in *different directions*.

[7] Every vector-valued function $\mathbf{f} : \mathbb{R}^n \to \mathbb{R}^m$, is an application such that $f_i : \mathbf{x} \mapsto f_i(\mathbf{x})$, for all $i = 1, 2, ..., m$. Such a function is said to be a *linear transformation* if, given

Definition 2.3.11 *For a family* $\mathfrak{F} = \{\mathbf{F_1}, \mathbf{F_2}, ..., \mathbf{F_k}\}$ *of Lipschitz mappings with contraction ratios* $\lambda_1, \lambda_2, ..., \lambda_k$*, we say that a non-empty compact set* A *is a* self-affine set *if it is invariant under* $\mathfrak{F}$ *and* $\mathbf{F_i}$ *are affine transformations, for all* $i = 1, 2, ..., k$.

It would be natural to look for a formula for the dimension of self-affine sets that generalizes Theorem 2.3.6, which is about self-similar sets. We would hope that the dimension depends on the affine transformation in a reasonably simple way, easily linked with the matrices and vectors that represent the affine transformation. Unfortunately, the situation is much more complex than this and we are not interested in going into this problem in more depth (for a extensive dissertation see [14]).

2.4 Fractal dimension of SBM, FBM and LSM

In this section we will apply the previous results on Fractal Geometry to the stochastic processes we had introduced in the first chapter. As a matter of fact, in Section 1.11 we stated that Standard Brownian Motions (SBM), Fractional Brownian Motions (FBM) and Lévy Stable Motions (LSM) with $\alpha \in (1, 2]$ possess a fractal structure. Moreover, in Section 1.3 we described self-affine stochastic processes and, in different part of the previous chapter, we proved that all those processes are self-affine.

$\mathbf{a_1}, \mathbf{a_2}, ..., \mathbf{a_m}$ vectors of $\mathbb{R}^n$, it is given by

$$\mathbf{f}(\mathbf{x}) = \begin{pmatrix} < \mathbf{a_1}, \mathbf{x} > \\ < \mathbf{a_2}, \mathbf{x} > \\ \vdots \\ < \mathbf{a_m}, \mathbf{x} > \end{pmatrix}.$$

If we indicate with A the matrix $m \times n$ obtained putting the vectors $\mathbf{a_1}, \mathbf{a_2}, ..., \mathbf{a_m}$ in columns

$$\mathrm{A} = \begin{pmatrix} \mathbf{a_1} \\ \mathbf{a_2} \\ \vdots \\ \mathbf{a_m} \end{pmatrix} = \begin{pmatrix} a_{1,1} & a_{1,2} & \dots & a_{1,n} \\ a_{2,1} & a_{2,2} & \dots & a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m,1} & a_{m,2} & \dots & a_{m,n} \end{pmatrix}$$

we can represent the m functions defined as below as an application $\mathbf{f} : \mathbb{R}^n \to \mathbb{R}^m$ such that

$$\mathbf{f} : \mathbf{T}(\mathbf{x}) = \mathrm{A}\mathbf{x}$$

where $\mathrm{A}\mathbf{x}$ represents the vector of $\mathbb{R}^m$ given by the matrix product between the matrix A $(m \times n)$ and the vector $\mathbf{x}$ $(n \times 1)$.

Moreover, since similarity are a subclass of affine transformations, it is trivial to prove that a similarity is an affinity such that

$$\det(\mathrm{A}) \neq 0\,.$$

Here, we will analyse the subject from the perspective of Fractal Geometry. First of all, we give an alternative characterization of self-affine processes, based on the concept of affinity introduced in the previous section. However, to do this we need the preliminary notion of *self-affine function.*

Definition 2.4.1 *A function* $\mathbf{g} : \mathbb{R}^k \to \mathbb{R}^h$ *is said to be a* self-affine function *if*

$$\|\mathbf{g}(a\mathbf{x}) - \mathbf{g}(a\mathbf{y})\| = \lambda \cdot \|\mathbf{g}(\mathbf{x}) - \mathbf{g}(\mathbf{y})\| . \tag{2.20}$$

for two positive real number a *and* λ*. If* $\lambda = a$*, we call that function a* self-similar function.

Note that functions which are invariant under scaling (or homogeneous functions) are also self-affine.

The concept of self-affine function can be extended to a stochastic process in the following way.

Definition 2.4.2 *Given a probability space* $(\Omega, \mathcal{F}, \mathbf{P})$ *and a set* $T \in [0, +\infty)$*, we say that a real-value stochastic process*

$$X(t) : \Omega \to \mathbb{R}^n \qquad \forall t \in T$$

is self-affine if[8]

$$\|X(a \cdot t) - X(a \cdot s)\| \overset{d}{\sim} \lambda \cdot \|X(t) - X(s)\| \tag{2.21}$$

for all $t, s \in T$*.*

If a process $\{X(t)\}_{t \in T}$ is a single-valued process (that is $X(t) \in \mathbb{R}$, for all $t \in T$). The norm can be replaced by the absolute value. Then

$$|X(a \cdot t) - X(a \cdot s)| \overset{d}{\sim} \lambda \cdot |X(t) - X(s)| .$$

Since the three stochastic processes are self-affine (in the sense of Definition A.3.1), they respect this additional relation. In fact, given $\{B(t)\}_{t \in T}$ a SMB, $\{B_H(t)\}_{t \in T}$ a FBM, and $\{M(t)\}_{t \in T}$ a LSM, we have

$$|B(a \cdot t) - B(a \cdot s)| \overset{d}{\sim} a^{\frac{1}{2}} \cdot |B(t) - B(s)|$$

$$|B_H(a \cdot t) - B_H(a \cdot s)| \overset{d}{\sim} a^H \cdot |B_H(t) - B_H(s)|$$

$$|M(a \cdot t) - M(a \cdot s)| \overset{d}{\sim} a^{\frac{1}{\alpha}} \cdot |M(t) - M(s)|$$

with $H \in (0, 1)$ and $\alpha \in (1, 2]$.

Having proved that they are self-similar, there is nothing else for us to do but showing that their graphs are fractals. We will prove this by computing the Hausdorff dimension of them, and thus using Definition 2.2.5. However, we need some preliminary definitions and results.

[8] Even if the stochastic process is a vector of $\mathbb{R}^n$, we do not use the bold type.

Definition 2.4.3 *Let $f : [a,b] \to \mathbb{R}$ be a function. We call* graph *of the function f the set given by the pair of the value of the independent variable and its image, that is*

$$\text{graph}_f([a,b]) := \left\{(x, f(x)) \in \mathbb{R}^2 : x \in [a,b]\right\} . \tag{2.22}$$

It can be proved that $\text{graph}_f([a,b])$ has dimension 1 (in the sense that both its Hausdorff dimension and topological dimension are equal to 1), if one of the following conditions hold:

- f has a continuous derivative, that is

$$f \in C^1([a,b]) ;$$

- f is of bounded variation, that is, given a partition $\Pi = \{x_0, x_1, ..., x_n\}$, with $a = x_0$ and $b = x_n$

$$\sup_{\Pi} \left\{ \sum_{i=1}^{n} |f(x_i) - f(x_{i-1})| \right\} < \infty .$$

However, it is possible for a continuous function to be sufficiently irregular to have a graph of dimension strictly grater than 1. Moreover, we have already seen that SBM, FBM and LSM are all almost surely nowhere differentiable and do not have bounded variations. This fact can be considered as the first heuristic evidence of the fractal structure of those processes.

We must remark that, in the literature on SBM, FBM or LSM, the confusion on this aspect reigns supreme. As a matter of fact, they are addressed almost every times as self-similar processes instead of being called self-affine. This is due to the fact that often there has been a slight misunderstanding about what (self-)similarity and (self-)affinity geometrically stand for. Roughly speaking, we have seen that similarities scale objects with the same ratio in all the directions; on the other hand, affinity allows to have different ratios in different direction (thus the object can be deformed). Regarding the three processes we have

$$a^{\frac{1}{2}} \cdot B(t) \overset{d}{\sim} B(a \cdot t)$$

$$a^{H} \cdot B_H(t) \overset{d}{\sim} B_H(a \cdot t)$$

and

$$a^{\frac{1}{\alpha}} \cdot M(t) \overset{d}{\sim} M(a \cdot t)$$

meaning that changing the temporal scale by a factor a and the spatial scale by a factor $a^{\frac{1}{2}}$ (or a^H, $a^{\frac{1}{\alpha}}$ respectively) gives a process indistinguishable from the original one: Thus all these processes have their *graphs* to be *self-affine*.

Self-similarity can be traced only in the following sense: The *trajectories* of SBM, FBM and LSM can be seen as self-similar applications, since in the spatial distribution of the trails

$$\{B(t)\}_{0\leq t\leq T} \overset{d}{\sim} a^{-\frac{1}{2}} \cdot \{B(t)\}_{0\leq t\leq a\cdot T}$$

$$\{B_H(t)\}_{0\leq t\leq T} \overset{d}{\sim} a^{-H} \cdot \{B_H(t)\}_{0\leq t\leq a\cdot T}$$

and

$$\{M(t)\}_{0\leq t\leq T} \overset{d}{\sim} a^{-\frac{1}{\alpha}} \cdot \{M(t)\}_{0\leq t\leq a\cdot T}\ ,$$

being indistinguishable except for a scale factor. However, this fact states that they have self-similar traits; they are *not* self-similar processes since they constitute a degenerate case of self-affine ones. Because of it, we will properly avoid to refer to them as self-similar processes.

Here, we state a very useful result (for its proof see [14]).

Proposition 2.4.4 *Let $f:[0,1]\to\mathbb{R}$ be a continuous β–Hölder function*

$$|f(t)-f(s)|\leq\lambda\cdot|t-s|^{\beta}$$

with $0<\beta\leq 1$ and $t,s\in[0,1]$. Then $\mathcal{H}^s\left[\mathrm{graph}_f([a,b])\right]<\infty$ and

$$\dim_{\mathcal{H}}\left[\mathrm{graph}_f([0,1])\right]\leq 2-\beta\,. \tag{2.23}$$

By σ-stability of Hausdorff dimension, the statements of the proposition remain true if $f:[0,+\infty)\to\mathbb{R}$ is only locally Hölder continuous. Thus this important result gives an upper bound for $\dim_{\mathcal{H}}\big[\mathrm{graph}_f([0,1])\big]$. A lower bound is harder to get. In order to do this, we need a more advanced result, taken from *Potential Theoretic Methods* (for its proof see again [14]).

Proposition 2.4.5 *Let $A\subset\mathbb{R}^n$ and a mass distribution*[9] *$\mathcal{M}$ on A. Then, given $q\geq 0$, for every $\delta>0$*

$$\mathcal{H}^q_\delta(A)\geq\frac{\mathcal{M}(A)^2}{\iint\limits_{\|\mathbf{x}-\mathbf{y}\|<\delta}\frac{\mathcal{M}(\mathrm{d}\mathbf{x})\,\mathcal{M}(\mathrm{d}\mathbf{y})}{\|\mathbf{x}-\mathbf{y}\|^q}}\,.$$

Hence, if $\mathcal{M}$ has finite q–energy *on A, that is*

$$\iint\limits_{\|\mathbf{x}-\mathbf{y}\|<\delta}\frac{\mathcal{M}(\mathrm{d}\mathbf{x})\,\mathcal{M}(\mathrm{d}\mathbf{y})}{\|\mathbf{x}-\mathbf{y}\|^q}<\infty$$

with $\mathbf{x},\mathbf{y}\in A$, then $\mathcal{H}^q(A)=\infty$ and, in particular,

$$\dim_{\mathcal{H}}(A)\geq q\,. \tag{2.24}$$

[9] Given a metric space X and a bounded set $A\subset X$, a measure $\mathcal{M}$ on A for which $0<\mathcal{M}(A)<\infty$ is called *mass distribution*.

The nowhere differentiability of Brownian motion established in the first chapter suggests that its graph may have dimension greater than one. First of all, we will show that SBM is a Hölder continuous process, in order to make Proposition 2.4.4 applicable.

Proposition 2.4.6 *Given $0 < \beta < \frac{1}{2}$ then, almost surely, a standard Brownian motion $\{B(t)\}_{t\in[0,1]}$ is everywhere locally β–Hölder continuous. That is $B : [0,1] \to \mathbb{R}$ satisfies*

$$|B(t) - B(s)| \leq \lambda \cdot |t-s|^{\beta} \qquad \text{a.s.} \tag{2.25}$$

where λ depends only on β.

Proof: By virtue of Definition 1.1.1, we have

$$\Pr\left\{|B(t) - B(s)| > |t-s|^{\beta}\right\} = \frac{1}{\sqrt{2\cdot\pi\cdot\mathbb{V}[B(t)-B(s)]}} \int_{|t-s|^{\beta}}^{+\infty} e^{-\frac{\{x-\mathbb{E}[B(t)-B(s)]\}^2}{2\cdot\mathbb{V}[B(t)-B(s)]}}\,\mathrm{d}x = \\ = \frac{1}{\sqrt{2\cdot\pi\cdot|t-s|}} \int_{|t-s|^{\beta}}^{+\infty} e^{-\frac{x^2}{2\cdot|t-s|}}\,\mathrm{d}x\,.$$

After a substitution $x = u\cdot\sqrt{|t-s|}$, we have[10]

$$\Pr\left\{|B(t) - B(s)| > |t-s|^{\beta}\right\} = \frac{1}{\sqrt{2\cdot\pi\cdot|t-s|}} \int_{|t-s|^{\beta-\frac{1}{2}}}^{+\infty} e^{-\frac{u^2}{2}}\cdot\sqrt{|t-s|}\,\mathrm{d}u = \\ = \frac{1}{\sqrt{2\cdot\pi}} \int_{|t-s|^{\beta-\frac{1}{2}}}^{+\infty} e^{-\frac{u^2}{2}}\,\mathrm{d}u\,.$$

Thus, it follows

$$\frac{1}{\sqrt{2\cdot\pi}} \int_{|t-s|^{\beta-\frac{1}{2}}}^{+\infty} e^{-\frac{u^2}{2}}\,\mathrm{d}u \;\leq\; c_1 \cdot \int_{|t-s|^{\beta-\frac{1}{2}}}^{+\infty} e^{-u}\,\mathrm{d}u$$

$$\frac{1}{\sqrt{2\cdot\pi}} \int_{|t-s|^{\beta-\frac{1}{2}}}^{+\infty} e^{-\frac{u^2}{2}}\,\mathrm{d}u \;\leq\; c_1 \cdot e^{-|t-s|^{\beta-\frac{1}{2}}}$$

$$\frac{1}{\sqrt{2\cdot\pi}} \int_{|t-s|^{\beta-\frac{1}{2}}}^{+\infty} e^{-\frac{u^2}{2}}\,\mathrm{d}u \;\leq\; c_1 \cdot e^{-\left(\beta-\frac{1}{2}\right)\cdot|t-s|}$$

for a constant c_1 which does not depend on t or s. Furthermore, since $-\frac{1}{2} \leq \beta - \frac{1}{2} \leq 0$, we also have

$$\frac{1}{\sqrt{2\cdot\pi}} \int_{|t-s|^{\beta-\frac{1}{2}}}^{+\infty} e^{-\frac{u^2}{2}}\,\mathrm{d}u \;\leq\; c_1 \cdot e^{-\left(\beta-\frac{1}{2}\right)\cdot|t-s|} \leq c_2 \cdot |t-s|$$

[10] Setting $x = f(u) = u\cdot\sqrt{|t-s|}$, it follows

$$x > |t-s|^{\beta} \quad\to\quad u\cdot\sqrt{|t-s|} > |t-s|^{\beta} \quad\to\quad u > |t-s|^{\beta-\frac{1}{2}}$$

and

$$\mathrm{d}x = f'(u)\,\mathrm{d}u = \sqrt{|t-s|}\,\mathrm{d}u\,.$$

for both constants c_1 and c_2 which do not depend on t or s.

Taking the binary intervals $\left[(m-1)\cdot 2^{-j}, m\cdot 2^{-j}\right]$, with $1 \leq m \leq 2^j$ (hence the binary interval are subsets of $[0,1]$), we have for some $j \geq k$:[11]

$$\Pr\left\{\left|B\left[(m-1)\cdot 2^{-j}\right] - B\left(m\cdot 2^{-j}\right)\right| > \left(2^{-j}\right)^{\beta}\right\} \leq c_2 \cdot \sum_{j=k}^{\infty} 2^{-j} = c_2 \cdot 2^{1-k}\,.$$

Hence there is an integer N such that, for al $j > N$, it follows

$$\Pr\left\{\left|B\left[(m-1)\cdot 2^{-j}\right] - B\left(m\cdot 2^{-j}\right)\right| \leq 2^{-j\cdot\beta}\right\} = 1\,.$$

If $|t-s| < 2^{-N}$, the interval $[t,s]$ (or $[s,t]$) may be expressed, except possibly for the endpoints, as a countable union of contiguous binary intervals of the form $\left[(m-1)\cdot 2^{-k}, m\cdot 2^{-k}\right]$ with $|t-s| \geq 2^{-k}$ and with no more than two intervals of any one length. Then, using the continuity of $\{B(t)\}$, we have $2^{-(k+1)} \leq |t-s|$, and hence almost surely

$$|B(t) - B(s)| \leq \sum_{j=k+1}^{\infty} 2^{-j\cdot\beta} = \frac{2^{-(k+1)\cdot\beta}}{1-2^{-\beta}} = \frac{2}{1-2^{-\beta}}\cdot 2^{-k\cdot\beta} \leq \frac{2}{1-2^{-\beta}}\cdot|t-s|^{\beta}$$

Setting $\frac{2}{1-2^{-\beta}} = \lambda$, the thesis follows

$$|B(t) - B(s)| \leq \lambda\cdot|t-s|^{\beta} \qquad \text{a.s.}$$

□

Now we are able to prove this crucial Theorem.

Theorem 2.4.7 *The graph of a standard Brownian motion $\{B(t)\}_{t\in[0,1]}$ has almost surely Hausdorff dimension equal to*

$$\dim_{\mathcal{H}}\left[\text{graph}_B([0,1])\right] = \frac{3}{2}\,. \tag{2.26}$$

Proof: By virtue of Theorem 2.4.4, since the SBM is almost surely β–Hölder continuous if and if only $0 < \beta < \frac{1}{2}$, we have

$$\dim_{\mathcal{H}}\left[\text{graph}_B([0,1])\right] \leq \frac{3}{2}\,. \tag{2.27}$$

For the lower estimate, we use the potential theoretic method. To get a lower bound on the dimension from this method it suffices to show finiteness

[11] The geometric series $\sum_{j=k}^{\infty}\left(\frac{1}{2}\right)^j$ converges to $\left(\frac{1}{2}\right)^{k-1}$ since $\frac{1}{2} < 1$.

of a single integral. In particular, in order to show for a random set A that $\dim_{\mathcal{H}}(A) \geq q$ almost surely, it suffices to show that

$$\mathbb{E}\left(\iint\limits_{\|\mathbf{x}-\mathbf{y}\|<\delta} \frac{\mathcal{M}(\mathrm{d}\mathbf{x})\,\mathcal{M}(\mathrm{d}\mathbf{y})}{\|\mathbf{x}-\mathbf{y}\|^q} \right) < \infty$$

for every $\delta > 0$ and $\mathbf{x}, \mathbf{y} \in A$, for a (random) measure on A (for the proof of this statement, see [32]).

There is a natural way of defining a mass distribution $\mathcal{M}$ on the graph of the function $B(t)$ by

$$\mathcal{M}(A) = \mathcal{L}^1\left(\left\{(t, B(t)) \in A : t \in [0,1]\right\}\right)$$

for $A \subset [0,1] \times \mathbb{R}$ Borel set, where $\mathcal{L}^1$ is the Lebesgue measure on $\mathbb{R}$. Then, for all Borel sets $A \subset \mathbb{R}^2$, Fubini's theorem allows us to write

$$\int_{\mathbb{R}^2} f(\mathbf{x})\,\mathcal{M}(\mathrm{d}\mathbf{x}) = \int_0^1 f\left(t, B(t)\right)\,\mathrm{d}t$$

for all bounded measurable functions f. Setting $f(\mathbf{x}) = \|\mathbf{x}\|^q$ and changing variable such that $\|\mathbf{x}-\mathbf{y}\|^2 = |B(t)-B(s)|^2 + |t-s|^2$ (Pythagoras' Theorem), the q–energy of $\mathcal{M}$ on A is given by

$$\begin{aligned}\iint\limits_{\|\mathbf{x}-\mathbf{y}\|<\delta} \frac{\mathcal{M}(\mathrm{d}\mathbf{x})\,\mathcal{M}(\mathrm{d}\mathbf{y})}{\|\mathbf{x}-\mathbf{y}\|^q} &= \int_0^1\int_0^1 \frac{\mathrm{d}t\,\mathrm{d}s}{\left(|B(t)-B(s)|^2 + |t-s|^2\right)^{\frac{q}{2}}} = \\ &= \int_0^1\int_0^1 \left(|B(t)-B(s)|^2 + |t-s|^2\right)^{-\frac{q}{2}}\,\mathrm{d}t\,\mathrm{d}s\end{aligned}$$

Applying Fubini's theorem, the expected value of such quantity can also be written as

$$\begin{aligned}&\mathbb{E}\left[\int_0^1\int_0^1 \left(|B(t)-B(s)|^2 + |t-s|^2\right)^{-\frac{q}{2}}\,\mathrm{d}t\,\mathrm{d}s\right] = \\ &= \int_0^1\int_0^1 \mathbb{E}\left[\left(|B(t)-B(s)|^2 + |t-s|^2\right)^{-\frac{q}{2}}\right]\,\mathrm{d}t\,\mathrm{d}s\,.\end{aligned}$$

Bounding the integrand (due to monotonicity) and , we get that

$$\int_0^1\int_0^1 \mathbb{E}\left[\left(|B(t)-B(s)|^2 + |t-s|^2\right)^{-\frac{q}{2}}\right]\,\mathrm{d}t\,\mathrm{d}s \leq 2\cdot\int_0^1 \mathbb{E}\left[\left(B(t)^2 + t^2\right)^{-\frac{q}{2}}\right]\,\mathrm{d}t\,.$$

Hence, if the latter expected value is bounded, the q–energy is also finite. Let we take into account this[12]:

$$\begin{aligned}\mathbb{E}\left[\left(B(t)^2+t^2\right)^{-\frac{q}{2}}\right] &= \frac{1}{\sqrt{2\cdot\pi\cdot t}}\int_0^{+\infty}\left(x^2+t^2\right)^{-\frac{q}{2}}\cdot e^{-\frac{x^2}{2\cdot t}}\,\mathrm{d}x = \\ &= \frac{1}{\sqrt{2\cdot\pi\cdot t}}\int_0^{+\infty}\left(t\cdot u^2+t^2\right)^{-\frac{q}{2}}\cdot e^{-\frac{u^2\cdot t}{2\cdot t}}\cdot\sqrt{t}\,\mathrm{d}u = \\ &= \frac{1}{\sqrt{2\cdot\pi}}\int_0^{+\infty}\left(t\cdot u^2+t^2\right)^{-\frac{q}{2}}\cdot e^{-\frac{u^2}{2}}\,\mathrm{d}u\,.\end{aligned}$$

Comparing the size of the summands in the integration suggests separating them at $u=\sqrt{t}$, that is

$$\int_0^{+\infty}\left(t\cdot u^2+t^2\right)^{-\frac{q}{2}}\cdot e^{-\frac{u}{2}}\,\mathrm{d}u = \int_0^{\sqrt{t}}\left(t\cdot u^2+t^2\right)^{-\frac{q}{2}}\cdot e^{-\frac{u}{2}}\,\mathrm{d}u + \int_{\sqrt{t}}^{+\infty}\left(t\cdot u^2+t^2\right)^{-\frac{q}{2}}\cdot e^{-\frac{u}{2}}\,\mathrm{d}u\,.$$

Moreover, they can be bounded in the following way

$$\int_0^{\sqrt{t}}\left(t\cdot u^2+t^2\right)^{-\frac{q}{2}}\cdot e^{-\frac{u}{2}}\,\mathrm{d}u \le \int_0^{\sqrt{t}}\left(t^2\right)^{-\frac{q}{2}}\,\mathrm{d}u$$

and

$$\int_{\sqrt{t}}^{+\infty}\left(t\cdot u^2+t^2\right)^{-\frac{q}{2}}\cdot e^{-\frac{u}{2}}\,\mathrm{d}u \le \int_{\sqrt{t}}^{+\infty}\left(t\cdot u^2\right)^{-\frac{q}{2}}\,\mathrm{d}u\,.$$

Since the two integrals are respectively equal to

$$\int_0^{\sqrt{t}}\left(t^2\right)^{-\frac{q}{2}}\,\mathrm{d}u = \int_0^{\sqrt{t}}t^{-q}\,\mathrm{d}u = t^{-q}\cdot\int_0^{\sqrt{t}}\mathrm{d}u = t^{-q}\cdot\sqrt{t} = t^{\frac{1}{2}-q}$$

and (with $q>1$, otherwise the integral diverges)

$$\int_{\sqrt{t}}^{+\infty}\left(t\cdot u^2\right)^{-\frac{q}{2}}\,\mathrm{d}u = t^{-\frac{q}{2}}\int_{\sqrt{t}}^{+\infty}u^{-q}\,\mathrm{d}u = t^{-\frac{q}{2}}\cdot\left(-\frac{t^{\frac{1}{2}-\frac{q}{2}}}{1-q}\right) = \frac{t^{\frac{1}{2}-q}}{q-1}\,,$$

hence their sum is equal to

$$\int_0^{\sqrt{t}}\left(t^2\right)^{-\frac{q}{2}}\,\mathrm{d}u + \int_{\sqrt{t}}^{+\infty}\left(t\cdot u^2\right)^{-\frac{q}{2}}\,\mathrm{d}u = t^{\frac{1}{2}-q}+\frac{t^{\frac{1}{2}-q}}{q-1} = \frac{q}{q-1}\cdot t^{\frac{1}{2}-q}\,.$$

Eventually we can write

$$2\cdot\mathbb{E}\left[\left(B(t)^2+t^2\right)^{-\frac{q}{2}}\right] \le c_1\cdot\frac{q}{q-1}\cdot t^{\frac{1}{2}-q}$$

[12]To solve the integral, we are making the substitution $x=f(u)=u\cdot\sqrt{t}$. Hence, it follows

$$x>0 \quad\to\quad u\cdot\sqrt{t}>0 \quad\to\quad u>0$$

and

$$\mathrm{d}x = f'(u)\,\mathrm{d}u = \sqrt{t}\,\mathrm{d}u\,.$$

for a constant c_1 which does not depend on t. Because of this, the integral of the expected value becomes

$$\int_0^1 \int_0^1 \mathbb{E}\left[\left(|B(t)-B(s)|^2+|t-s|^2\right)^{-\frac{q}{2}}\right] \mathrm{d}t\,\mathrm{d}s \leq 2\cdot\int_0^1 \mathbb{E}\left[\left(B(t)^2+t^2\right)^{-\frac{q}{2}}\right]\mathrm{d}t$$
$$\leq c_1\cdot\frac{q}{q-1}\cdot\int_0^1 t^{\frac{1}{2}-q}\,\mathrm{d}t$$

Since

$$\int_0^1 t^{\frac{1}{2}-q}\,\mathrm{d}t = \left(\frac{t^{\frac{3}{2}-q}}{\frac{3}{2}-q}\right)\Bigg|_0^1 < \infty$$

if and only if $q<\frac{3}{2}$, by virtue of Proposition 2.4.5, we have

$$\dim_{\mathcal{H}}\left[\mathrm{graph}_B([0,1])\right] \geq \frac{3}{2}\,. \tag{2.28}$$

Together with (2.27), it necessarily follows

$$\dim_{\mathcal{H}}\left[\mathrm{graph}_B([0,1])\right] = \frac{3}{2}\,.$$

□

Regarding FBM and LSM, we have seen that both have unbounded variation and are almost surely nowhere differentiable. This fact allows us to believe that they are both fractals, having a graph dimension grater than one. As a matter of fact, using rather more sophisticated techniques from Probability Theory, the following results hold true (for a sketch of proof see [14]).

Proposition 2.4.8 *Given $0<\beta<H$ then, almost surely, a fractional Brownian motion $\{B_H(t)\}_{t\in[0,1]}$ is everywhere locally β–Hölder continuous. That is $B_H:[0,1]\to\mathbb{R}$ satisfies*

$$|B_H(t)-B_H(s)| \leq \lambda\cdot|t-s|^{\beta} \qquad \text{a.s.} \tag{2.29}$$

where λ depends only on β.

Hence a theorem on FBM's graph dimension follows like in the SBM case.

Theorem 2.4.9 *The graph of a fractional Brownian motion $\{B_H(t)\}_{t\in[0,1]}$ has almost surely Hausdorff dimension equal to*

$$\dim_{\mathcal{H}}\left[\mathrm{graph}_{B_H}([0,1])\right] = 2-H\,, \tag{2.30}$$

where H is the Hurst exponent of the process.

About LSM, a corresponding β–Hölder continuity does not exist (let we remark that LSM are almost surely discontinuous, unless $\alpha = 2$). However, we can state the following result about graph dimension.

Theorem 2.4.10 *The graph of a Lévy stable motion* $\{M(t)\}_{t\in[0,1]}$ *has almost surely Hausdorff dimension equal to*

$$\dim_{\mathcal{H}}\big[\mathrm{graph}_M([0,1])\big] = \max\left\{1, 2-\frac{1}{\alpha}\right\}, \tag{2.31}$$

where α *is the shape parameter of the process.*

Note that, the latter result describes the same occurrence of Theorem 1.6.12. As a matter of fact, if $0 < \alpha \leq 1$, then

$$\max\left\{1, 2-\frac{1}{\alpha}\right\} = 1$$

since $2 - \frac{1}{\alpha} \leq 1$. On the other hand, if if $1 < \alpha \leq 2$, then

$$\max\left\{1, 2-\frac{1}{\alpha}\right\} = 2-\frac{1}{\alpha} > 1\,.$$

Thus, if and if only $\alpha \in (1, 2]$, the LSM is a self-affine process (see Theorem 1.6.12) and its graph has a Hausdorff dimension grater than one, being so a fractal.

Note that, both Theorem 2.4.9 and 2.4.10 lead to the same result of Theorem 2.4.7 considering, respectively, a fractional Brownian motion with Hurst exponent $H = \frac{1}{2}$ and a Lévy stable motion with shape parameter $\alpha = 2$.

2.5 Multifractal measures

A mass distribution $\mathcal{M}$ may extend itself over a region of the hyperspace in very different ways, sometimes with concentration areas which are distributed in a highly irregular fashion. Let we take into account an open n-ball where the mass locally follows a power law with α index, that is

$$\mathcal{M}\big[B(\mathbf{x}, r)\big] \approx r^{\alpha}$$

for small values of r. We will see that such a definition implies interesting features, such as giving different fractals for different values of α: In this way, a whole range of fractals may arise from a single measure. A measure $\mathcal{M}$ with such a rich behaviour is named a *multifractal measure* (or just *multifractal*). For the same reasons about fractals, a precise definition will be avoided since it might cause improper exclusions.

Multifractal measures have been recognized in several different fields, such as fluid turbulence, earthquake and rainfall distributions, neural networks and share prices as well. The next chapter will analyze a model proposed by Mandelbrot and two former students of him which use a multifractal measure in order to give "different weights" at different stages of the trading time, hence influencing price changes. Here, we shall examine the structure of such multifractals and their inter-relationships.

As we have done for fractal introduction, there are two basic approaches to multifractal analysis:

- *Fine Theory*, where the structure and dimensions of fractals are examined as they arise themselves. In many ways, fine multifractal analysis is linked to finding Hausdorff dimension of sets;
- *Coarse Theory*, where we consider the distributional irregularities of the measure of n-balls of small positive radius, taking the limit as $r \to 0$. This approach is more intuitive and it is similar to the one used to introduce similarity dimension.

There are many parallels between the two theories, and for many measures both approaches lead to the same "multifractal spectra" (see Definition 2.5.1 and 2.5.4). As we have done with similarity dimension, we begin with the coarse theory rather than with the fine one.

Let we consider a r-mesh hyper-cube in $\mathbb{R}^n$ that is a hyper-cube C of the form

$$C := [m_1 \cdot r, (m_1 + 1) \cdot r] \times [m_2 \cdot r, (m_2 + 1) \cdot r] \times ... \times [m_n \cdot r, (m_n + 1) \cdot r]$$

where r is a positive real number and $m_1, m_2, \ldots, m_n \in \mathbb{Z}$. For a finite measure $\mathcal{M}$ on $\mathbb{R}^n$, we write

$$N_r(\alpha) := \#\{r\text{-mesh hyper-cubes } C : \mathcal{M}(C) \geq r^\alpha\} \tag{2.32}$$

with $\alpha \geq 0$.

Definition 2.5.1 *We define* coarse multifractal spectrum[13] *of* $\mathcal{M}$ *as*

$$f_C(\alpha) := \lim_{\varepsilon \to 0} \lim_{r \to 0} \frac{\max\{\log[N_r(\alpha + \varepsilon) - N_r(\alpha - \varepsilon)], 0\}}{\log\left(\frac{1}{r}\right)} \tag{2.33}$$

for $\alpha \geq 0$. *To allow for the eventuality of the limit not to exist, we define the* lower *and the* upper coarse multifractal spectrum *of* $\mathcal{M}$ *as*

$$\underline{f}_C(\alpha) := \lim_{\varepsilon \to 0} \liminf_{r \to 0} \frac{\max\{\log[N_r(\alpha + \varepsilon) - N_r(\alpha - \varepsilon)], 0\}}{\log\left(\frac{1}{r}\right)} \tag{2.34}$$

[13] Note the analogy with the similarity dimension.

and

$$\overline{f}_C(\alpha) := \lim_{\varepsilon\to 0} \limsup_{r\to 0} \frac{\max\{\log\left[N_r(\alpha+\varepsilon) - N_r(\alpha-\varepsilon)\right], 0\}}{\log\left(\frac{1}{r}\right)} \tag{2.35}$$

for $\alpha \geq 0$.

The definition formula (2.33) implies that, given $\eta > 0$ and $\varepsilon > 0$ small enough, then

$$r^{-f_C(\alpha)+\eta} \leq N_r(\alpha+\varepsilon) - N_r(\alpha-\varepsilon) \leq r^{-f_C(\alpha)-\eta}$$

for sufficiently small r. Alternatively speaking, $-f_C(\alpha)$ is the power law exponent for the number of r-mesh hyper-cubes C such that

$$\mathcal{M}(C) \sim r^{\alpha}$$

when $r \to 0$. Moreover, we used the function $\max\{\,\cdot\,, 0\}$ into the definition above in order to exclude negative values of the multifractal spectra.

Coarse multifractal spectra are often awkward to estimate. However, there exists a relation that allows to get an upper limit and sometimes also an equality (and hence a different way to compute it). Let we introduce the *q-th moment sum of $\mathcal{M}$*.

Definition 2.5.2 *Given the set $\mathfrak{M}_r$ of all the r-mesh hyper-cubes C such that $\mathcal{M}(C) > 0$, we define the* q-th moment sum of $\mathcal{M}$ *as*

$$m_r(q) := \sum_{\mathfrak{M}_r} \mathcal{M}(C)^q \tag{2.36}$$

for $q \in \mathbb{R}$ and $r > 0$.

It is important to identify the power law behaviour of $m_r(q)$ throughout the following function

$$\beta(q) := \lim_{r\to 0} \frac{\log\left[m_r(q)\right]}{\log\left(\frac{1}{r}\right)}, \tag{2.37}$$

assuming that this limit exists. It is quite unexpected that those moment sums are related to the $N_r(\alpha)$. But, using (2.32), it follows that, if $q \geq 0$ and $\alpha \geq 0$[14]

$$m_r(q) = \sum_{\mathfrak{M}_r} \mathcal{M}(C)^q \geq r^{q\cdot\alpha} \cdot N_r(\alpha)$$

[14]Since $\mathcal{M}(C) \geq r^{\alpha}$, hence $\mathcal{M}(C)^q \geq r^{q\cdot\alpha}$ being q a positive number. Subsequently, we have $\#\left(\mathfrak{M}_r\right) = \#\{r\text{-mesh hyper-cubes } C : \mathcal{M}(C)^q \geq r^{q\cdot\alpha}\}$, which leads to the first inequality. On the other hand, if q is negative $\mathcal{M}(C)^q \leq r^{q\cdot\alpha}$, we have $\#\left(\mathfrak{M}_r\right) = \#\{r\text{-mesh hyper-cubes } C : 0 < \mathcal{M}(C)^q \leq r^{q\cdot\alpha}\}$, for the second inequality.

and if $q < 0$

$$m_r(q) = \sum_{\mathfrak{M}_r} \mathcal{M}(C)^q \geq r^{q\cdot\alpha} \cdot \#\left\{r\text{-mesh hyper-cubes } C : 0 < \mathcal{M}(C) \leq r^{\alpha}\right\}$$

These inequalities lead to a useful relationship between $f_C(\alpha)$ and $\beta(q)$ which is stated in the following proposition.

Proposition 2.5.3 *Let $\mathcal{M}$ be a finite measure on $\mathbb{R}^n$ and assume that the limit*

$$\beta(q) = \lim_{r\to 0} \frac{\log\left[m_r(q)\right]}{\log\left(\frac{1}{r}\right)} = \lim_{r\to 0} -\frac{\log\left[m_r(q)\right]}{\log\left(r\right)},$$

exists. Then, for all $\alpha \geq 0$,

$$\underline{f}_C(\alpha) \leq \overline{f}_C(\alpha) \leq \beta^{\star}(\alpha) \tag{2.38}$$

where $\beta^{\star} : \mathbb{R} \to \mathbb{R}$ is the Legendre transform[15] *of β, that is*

$$\beta^{\star}(\alpha) = \inf_{q\,\in\,\mathbb{R}} \left\{q \cdot \alpha + \beta(q)\right\}.$$

Proof: Firstly, let we consider $q \geq 0$. Then, given $\varepsilon > 0$, (2.35) implies that

$$m_r(q) \geq r^{q\cdot(\alpha+\varepsilon)} \cdot N_r(\alpha+\varepsilon) \geq r^{q\cdot(\alpha+\varepsilon)} \cdot r^{-\overline{f}_C(\alpha)+\varepsilon}$$

for arbitrary small r. It follows from (2.37) that

$$\begin{aligned}
\beta(q) = \lim_{r\to 0} -\frac{\log\left[m_r(q)\right]}{\log(r)} &\geq \lim_{r\to 0} -\frac{\log\left[r^{q\cdot(\alpha+\varepsilon)-\overline{f}_C(\alpha)+\varepsilon}\right]}{\log(r)} = \\
&= \lim_{r\to 0} -\frac{\frac{\left[q\cdot(\alpha+\varepsilon)-\overline{f}_C(\alpha)+\varepsilon\right]\cdot r^{q\cdot(\alpha+\varepsilon)-\overline{f}_C(\alpha)+\varepsilon-1}}{r^{q\cdot(\alpha+\varepsilon)-\overline{f}_C(\alpha)+\varepsilon}}}{\frac{1}{r}} = \\
&= \lim_{r\to 0} -\frac{\left[q\cdot(\alpha+\varepsilon)-\overline{f}_C(\alpha)+\varepsilon\right]\cdot\frac{1}{r}}{\frac{1}{r}} = \\
&= -\left[q\cdot(\alpha+\varepsilon)-\overline{f}_C(\alpha)+\varepsilon\right].
\end{aligned}$$

Hence

$$-\beta(q) \leq q\cdot(\alpha+\varepsilon) - \overline{f}_C(\alpha) + \varepsilon$$

and, letting $\varepsilon \to 0$

$$\overline{f}_C(\alpha) \leq q\cdot\alpha + \beta(q),.$$

Then, let we consider $q < 0$. Given $\varepsilon > 0$, (2.34) implies that

$$m_r(q) \geq r^{q\cdot(\alpha-\varepsilon)} \cdot r^{-\overline{f}_C(\alpha)-\varepsilon}$$

[15] If $\beta(q)$ is differentiable, its Legendre transform is the intersection of the tangent of $\beta(q)$ of slope equal to $-\alpha$ with the β-axis.

for arbitrary small r. It follows that

$$\beta(q) \geq \lim_{r\to 0} -\frac{\log\left[r^{q\cdot(\alpha-\varepsilon)-\overline{f}_C(\alpha)-\varepsilon}\right]}{\log(r)} = \lim_{r\to 0} -\frac{\left[q\cdot(\alpha-\varepsilon)-\overline{f}_C(\alpha)-\varepsilon\right]\cdot\frac{1}{r}}{\frac{1}{r}} = \\ = -\left[q\cdot(\alpha-\varepsilon)-\overline{f}_C(\alpha)-\varepsilon\right] .$$

Hence

$$-\beta(q) \leq q\cdot(\alpha-\varepsilon)-\overline{f}_C(\alpha)+\varepsilon$$

and, letting $\varepsilon \to 0$

$$\overline{f}_C(\alpha) \leq q\cdot\alpha+\beta(q) .$$

Since $\underline{f}_C(\alpha) \leq \overline{f}_C(\alpha)$ by definition, we have the thesis

$$\underline{f}_C(\alpha) \leq \overline{f}_C(\alpha) \leq \beta^\star(\alpha) .$$

□

Note that choosing an appropriate measure can lead to an equality relation, having a Legendre transform equal to the coarse multifractal spectrum.

On the other hand, fine multifractal analysis sees fractals as determined by their local *intensity* of a measure. Given a set $A \subset \mathbb{R}^n$, we are interested in studying the sets of points of A. For $\alpha \geq 0$, we define the following set:

$$F_\alpha = \left\{ \mathbf{x} \in A : \lim_{r\to 0} \frac{\log\big(\mathcal{M}\left[B(\mathbf{x},r)\right]\big)}{\log(r)} = \alpha \right\} \tag{2.39}$$

with $B(\mathbf{x},r) \subseteq A$. Thus F_α includes those points where the limit in (2.39) exists and is equal to α.

In fine multifractal analysis we aim at finding the Hausdorff dimension of F_α for a range of values of α. Thus we will focus on the *fine Hausdorff multifractal spectrum.*

Definition 2.5.4 *Let a set $A \subset \mathbb{R}^n$, $\alpha \in \mathbb{R}_0^*$, and a set F_α given by* (2.39)*. We define* fine Hausdorff multifractal spectrum *the Hausdorff dimension of F_α, that is*

$$f_{\mathcal{H}}(\alpha) := \dim_{\mathcal{H}}\left(F_\alpha\right) . \tag{2.40}$$

It can be proved that

$$0 \leq f_{\mathcal{H}}(\alpha) \leq \alpha . \tag{2.41}$$

An other important result we will prove is the one stated below, which gives a lower bound coarse multifractal spectrum, instead of Proposition 2.5.3, which establishes an upper bound.

Proposition 2.5.5 *Let $\mathcal{M}$ be a finite measure on $\mathbb{R}^n$. Then*

$$f_{\mathcal{H}}(\alpha) \leq \underline{f}_C(\alpha) \leq \overline{f}_C(\alpha) \tag{2.42}$$

Proof: The right-hand inequality is obvious. For convenience we will write f instead of $f_{\mathcal{H}}(\alpha)$.

Given $0 < \varepsilon < f$, then $\mathcal{H}^{f-\varepsilon}(F_\alpha) = \infty$. By (2.39) there is a set $F_\alpha^0 \subset F_\alpha \subset \mathbb{R}^n$ with $\mathcal{H}^{f-\varepsilon}(F_\alpha^0) > 1$. Moreover, by Egorov's Theorem, there exists a number $k_0 > 0$ such that

$$3^n \cdot r^{\alpha+\varepsilon} \leq \mathcal{M}\left[B(\mathbf{x}, r)\right] < \left(\frac{r}{2}\right)^{\alpha-\varepsilon}$$

for all $\mathbf{x} \in F_\alpha^0$ and $0 < r \leq k_0$[16]. We may choose δ, with $0 < \delta \leq \frac{k_0}{2}$, such that $\mathcal{H}_\delta^{f-\varepsilon}(F_\alpha^0) \geq 1$.

For each $r \leq \delta$, we consider a r-mesh hyper-cube $C \subset \mathbb{R}^n$ that intersect F_α^0. Such a set C contains a point $\mathbf{x}$ of F_α^0 with[17]

$$B(\mathbf{x}, r) \subset C \cup C_1 \cup C_2 \cup ... \cup C_{3^n-1} \subset B(\mathbf{x}, 2r\sqrt{n})$$

where C_i, with $i = 1, 2, \ldots, 3^n - 1$, are the r-mesh hyper-cubes immediately "around" C. By the previous inequality, it follows

$$3^n \cdot r^{\alpha+\varepsilon} \leq \mathcal{M}\left[B(\mathbf{x}, r)\right] \leq \mathcal{M}\left(C \cup C_1 \cup C_2 \cup ... \cup C_{3^n-1}\right) \leq \mathcal{M}\left[B(\mathbf{x}, 2r\sqrt{n})\right] < (r\sqrt{n})^{\alpha-\varepsilon}$$

so that

$$r^{\alpha+\varepsilon} \leq \mathcal{M}\left(C_0\right) < (r\sqrt{n})^{\alpha-\varepsilon},$$

where C_0 is one of $C, C_1, C_2, \ldots, C_{3^n-1}$. By definition of $\mathcal{H}_\delta^{f-\varepsilon}$ there are at least

$$(r\sqrt{n})^{\varepsilon-f} \cdot \mathcal{H}_\delta^{f-\varepsilon}(F_\alpha^0) \geq (r\sqrt{n})^{\varepsilon-f}$$

distinct r-mesh hyper-cubes that intersect F_α^0. So there are at least $\frac{(r\sqrt{n})^{\varepsilon-f}}{3^n-1}$ distinct r-mesh hyper-cubes C_0 that satisfy $r^{\alpha+\varepsilon} \leq \mathcal{M}\left(C_0\right) < (r\sqrt{n})^{\alpha-\varepsilon}$.

Hence, we conclude that, for $r \leq \delta$

$$N_{r\sqrt{n}}(\alpha+\varepsilon) - N_{r\sqrt{n}}(\alpha-\varepsilon) \geq \frac{1}{3^n} \cdot (r\sqrt{n})^{\varepsilon-f}.$$

So, using (2.34), we have

$$f - \varepsilon \leq \underline{f}_C(\alpha).$$

Letting $\varepsilon \to 0$, we have the thesis

$$f_{\mathcal{H}}(\alpha) = f \leq \underline{f}_C(\alpha) \leq \overline{f}_C(\alpha).$$

[16] It is easy to prove that $k_0 = e^{\frac{(\varepsilon-\alpha)\cdot\ln(2)-n\cdot\ln(3)}{2\varepsilon}}$.

[17] Note that $B(\mathbf{x}, r), C \cup C_1 \cup C_2 \cup ... \cup C_{3^n-1}$ and $B(\mathbf{x}, 2r\sqrt{n})$ as diameters equal to $2r, 3r\sqrt{n}$ and $4r\sqrt{n}$ respectively.

□

As for the three fractals we have introduced (the middle third Cantor set, the von Koch curve and the Sierpiński triangle), which have similarity dimension equal to Hausdorff dimension, many common multifractal measures have coarse multifractal spectrum equal to fine hausdorff multifractal spectrum. If this occurs, we have

$$f_{\mathcal{H}}(\alpha) = f_C(\alpha) = \beta^\star(\alpha)\,. \tag{2.43}$$

Since the first eventuality holds true if the fractal is self-similar, we could image that the previous relation is valid for *self-similar measures*. For its proof and for an in-depth dissertation of such measures, see [14]. The following definition will be used in the next chapter.

Definition 2.5.6 *Given a closed set $A \subset \mathbb{R}^n$ and a family of contracting similarities $\mathfrak{F} = \{\mathbf{F_1}, \mathbf{F_2}, ..., \mathbf{F_k}\}$ for which the the open set condition is fulfilled, a multifractal measure $\mathcal{M}$ is said to be a* self-similar measures *if*

$$\frac{\mathcal{M}[\mathbf{F_i}(A_1)]}{\mathcal{M}(A_1)} = \frac{\mathcal{M}[\mathbf{F_i}(A_2)]}{\mathcal{M}(A_2)} = \cdots = \frac{\mathcal{M}[\mathbf{F_i}(A_n)]}{\mathcal{M}(A_n)} \tag{2.44}$$

for every $i = 1, 2, \ldots, k$, where $A_1 \subseteq A_2 \subseteq ... \subseteq A_n$ is a non-decreasing sequence of compact sets contained in A.

The exposition about multifractals should have highlighted that the generalization from fractal sets to multifractal measures involves the passage from geometric objects characterized primarily by one number to geometric objects characterized primarily by a function (that is the measure itself). This function can also be a probability distribution, after being renormalized and plotted suitably. Moreover, the generalization for fractal sets to multifractal measures involves the passage from one fractal dimension to an infinite number of dimensions (related to the values of $f_{\mathcal{H}}(\alpha)$ or $f_C(\alpha)$, depending on the case).

Chapter 3

A different model to describe price fluctuations

The development of this chapter is based on the ideas of Mandelbrot about market misbehavior. He developed a completely different model based on fractals and multifractal measures together with two former students of his, Laurent Calvet and Adlai Fisher. The model firstly appeared in 1996 ([7], [8] and [9]). In 2003 it was developed by Mandelbrot himself in [37] and then implemented by the other two authors after Mandelbrot's death in [10].

We have seen that the peculiarity of FBM as a stochastic process for log-prices' description is mainly given by its fractal structure and long-range dependence. Moreover, the fact of being Gaussian bears further advantages. However, its law implies also that the oscillations of FBM at fine scales are uniform which comes as a disadvantage in various situations. Real market signals often possess an erratically changing oscillatory behavior which have earned them the name *multifractal processes*, but which also limits the appropriateness of FBM as a model. This rich structure at fine scales may serve as a valuable indicator, and ignoring it might mean to miss out on relevant information.

The model, namely a **Multifractal Model of Asset Returns** (MMAR), is anchored on the assertion that market time is *relative* in a certain sense. Financial markets would work according to an intrinsic "trading time", distinct from the linear physical time. That time accelerates the clock in high-volatility periods and slows down during those moments of placidity. In mathematical terms, we can write an equation showing the relationship between the two time structures and use it in order to generate the same irregularities of real financial prices. That phenomenon highlights a very important involvement already known by the most part of the financial practitioners. They often refers to a "fast" market and "slow" market on the strength of their volatility's perception in that moment.

Analogously, following a popular opinion, financial prices' patterns have

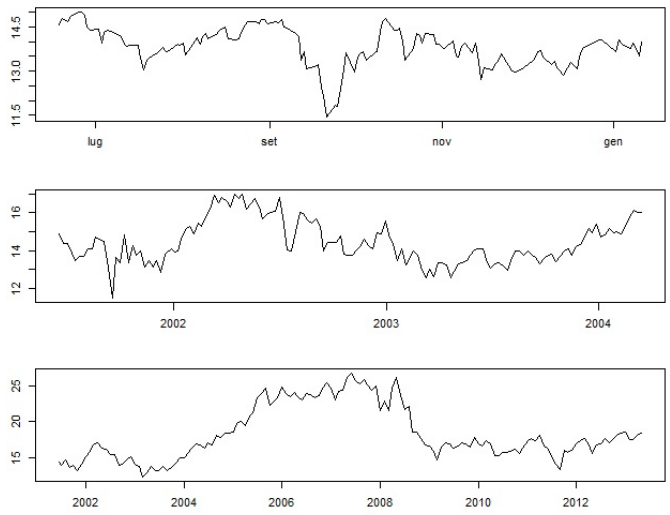


Figure 3.1: Which are the daily/weekly/mountly records?

all the same aspect: Without a caption about data's time-scale, none can state with certainty if the graph is taken by the last ten minutes, ten days, ten weeks and so long. Hence markets are *scale-invariance*. As already seen, this quality defines the charts as fractal curves and many powerful tools of mathematical analysis become available.

With regard to financial markets, scale-invariance means that the price evolution process can be described in terms of minutely, hourly, or daily recorded data, but the principle property of the process, like the distribution of the price variations, will always be of the same general form, with only a scale parameter that needs to be adjusted for a change of the time scale.

We have already seen in the second chapter that the technical term for such a form of close likeness between the parts and the whole is *self-affinity*. This property is related to the better-known concept of fractals named *self-similarity*, in which every feature of a picture is reduced or enlarged by the same ratio. Financial market charts, however, are far from being self-similar. Hence, the transformation from the whole to the parts must shrink the time scale (the horizontal axis) more than the price scale (the vertical axis). The geometric relation of the whole to its parts is said to be one of self-affinity. Of course, after the application of a suitable rescaling transformation, which takes the form of a contraction dependent upon the time scale only (an anisotropic transformation), a self-affinity in price variations may become self-similar. We will focus on this concept in the next section, introducing a tool named *cartoon* that, simplifying virtual price changes, will clarify that phenomenon.

However, before going on, we might wonder if we really need a further

model to describe price changes. Thus, are Brownian motions, fractional Brownian motions or Lévy-stable motions able to describe the behaviour of the financial markets? Unfortunately, they are not. Each one has got desirable features (scale-invariance, long-range dependence, heavy tails, discontinuities, volatility clustering), but no one possesses all.

In order to show this, we have taken into account more than 3.000 daily log-returns of the Italian stock ENI.MI[1] and compared with the realizations generated by the previous models. For the purpose of doing this, we worked on the standardized time-series of the log-returns.

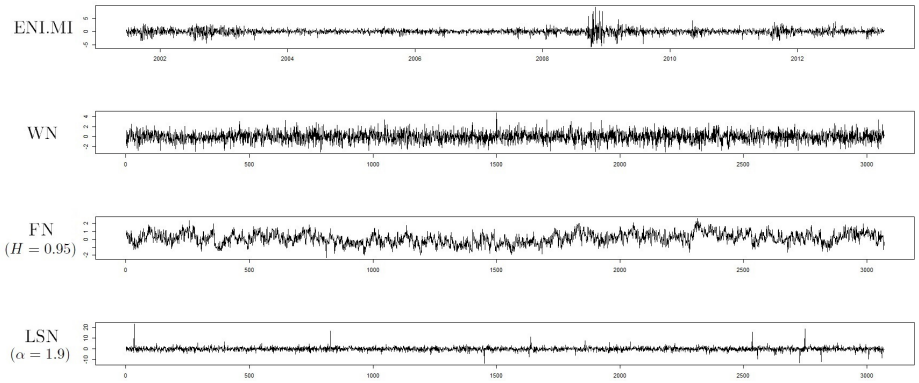


Figure 3.2: Who tells the truth?

Since prices are assumed to move as a BM, FBM or LSM, the corresponding log-return should follow a Gaussian white noise in the first case, and a Gaussian fractional white noise and a Lévy-stable noise[2] in the other two.

The figure shows evidently what we have just stated (see Figure 1 of the Introduction and Section 1.7). Even thought with different degrees of precision, none really fits reality. Of course, the Gaussian white noise is clearly the one which worse describes price changes. However, it is the most commonly used, since it represents the circumstance expected using a Brownian motion to model price movements (and hence the one used to make option pricing). Thus it is undeniable that we do need a more fitting and realistic model.

The MMAR aims at integrating all those features a financial time-series exhibits. As a matter of fact we will show that, starting from a stochastic

[1] We preferred to work with returns rather than prices since the inconsistency between the virtual records and the forecasts provided by theoretical models are more evident.

[2] These topics will be deeply analyzed in the Appendix D. For the moment, they can be simply thought as the *formal derivatives* of a FBM and a LSM respectively. Moreover LSM was taken with $\beta = 0$, being so symmetric.

process having long-range dependence (like a fractional Brownian motion with $\frac{1}{2} < H < 1$) and using a multifractal measure in order to create a compounded process, we will be able to get a new process which has both a long-range dependence and heavy tails. A a matter of fact, since a multifractal measure is an entity such that

$$\mathcal{M}\big[B(\mathbf{x}, r)\big] \approx r^{\alpha}\,,$$

thus involing a power-law, it suffices to create a power-law distribution (like Pareto's and Lévy-stable ones). Moreover, all the so-called multifractal (and fractal) formalism, deeply analyzed in the previous chapter, will turn out to be essential to understand the MMAR.

In the following, we will introduce two different multifractal approaches, both suggested by Mandelbrot. The first one is very intuitive and more practical; the second one is mainly theoretic and requires a considerable employment of mathematical tools.

3.1 A simple approach to a multifractal model

In this section we will introduce a very simple model to describe price changes. Because of its naiveté we are aware this model cannot describe price movements in a truthful fashion. However it will be able to catch scale-invariance and present the idea of multifractal applied to financial records.

Mandelbrot's ideas was to find a model with suitable unchanging properties, such as the invariance in price changes. In fact, despite the sense of statisticians, a scale-invariant model possesses relevant properties, such as the possibility to use it irrespective of the frequency of the data. Moreover, it should be a matter of common sense: Time frequencies surely give details at different scales, but it cannot be accepted the fact that, using daily records, a certain model is adopted, but using weekly prices we find a completely different one (strange but true, that is what happens using ARCH/GARCH models to describe volatility).

The starting point is very simple: We start drawing a simple chart that interpolates price changes from time $t_0 = 0$ to a later time $t_n = 1$ in successive steps. Next, the intermediate time intervals, $(t_0, t_1]$,$(t_1, t_2]$,...,$(t_{n-1}, t_n]$, may be chosen arbitrarily; they may represent a second, a hour, a day or a year.

The process begins with a price represented by a straight trend line called *initiator* (the blue line of Figure 3.3)[3]. Later, a broken line called *generator* is used to create the pattern that correspond to an up-and-down price movement. In Figure 3.3, the generator consists of three pieces (three is the least number necessary to create a non-degenerate plot) that are intersected along

[3]Figures of this section are taken from [24] and modified in some features.

the straight blue initiator. Then, each new straight line is used as initiator and the procedure made by the generator is reapplied recursively. Note that the interpolated generator is inverted for each descending piece. Repeating these steps we aim at describing a price curve. Moreover, at each stage the same form (the generator) is reproduced at increasingly compressed scales. Both the horizontal axis (time scale) an the vertical axis (price scale) are squeezed to fit the horizontal and vertical boundaries of each piece of generator. What we finally get is called *cartoon*. With this technique, using different generators, a widely variety of patterns can be represented. The final shape is very sensitive to the definition of the generator.

So, starting from a straight line which covers one entire period (we may say 1 year), interpolating the generator with the just-made new initiators, we are able to describe a self-affine structure at lower frequencies (at the first interpolation we obtain the record at the times $t_1 = 5$ mounths an 10 days, $t_2 = 6$ mounths and 20 days).

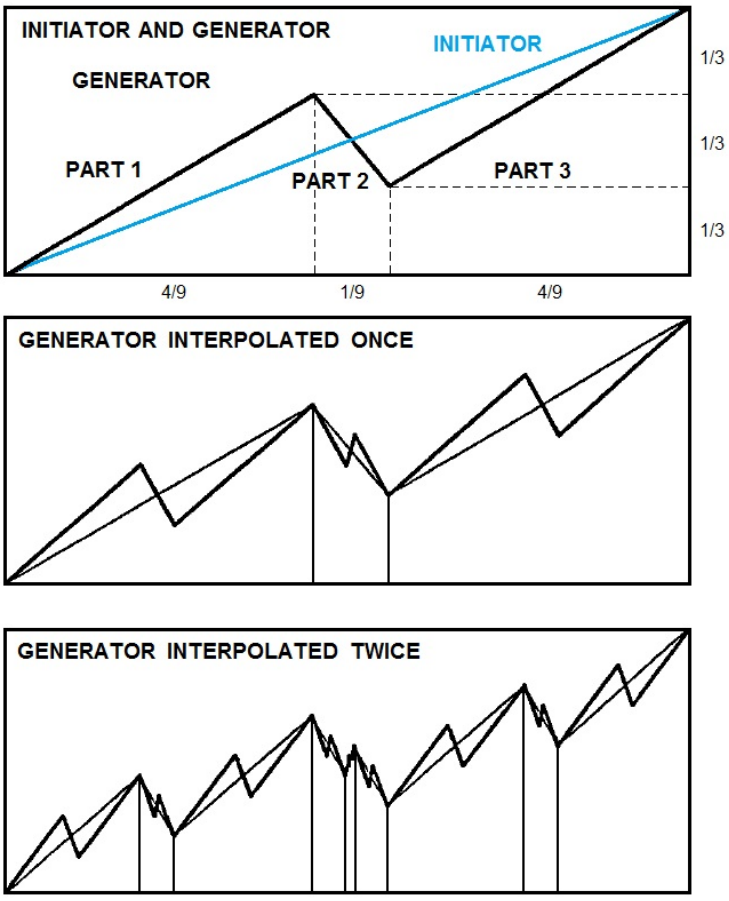


Figure 3.3: The "pseudo-Brownian" *cartoon*

Hence the same process continues and, in theory, it has no end. Of course in practice, it makes no sense to interpolate down to time intervals shorter than those between trading transactions (which may be of the order of a minute). The fact that each piece possesses a shape like the whole is not a surprise, since we built it in. The real surprise is that such a fractal curve has been built very easily and that the final outcome is plausible as price records during a period. This is the privilege in using fractal: Using a very simply defined curve, we can create a wealth of structure. Furthermore, the same model is observed at all time scales, leading so to a *general* structure.

Now we will discuss with more attention the exact form of the generator. As a matter of fact, the one of the figure could be addressed as a "pseudo-Brownian" cartoon. If we call as $\ln[S(t_i)]$ the (standard) log-price at time $t_i \in [0,1]$, we should see a particular regularity. Let we consider the first interpolation in Figure 3.3. We have $t_0 = 0$, $t_1 = \frac{4}{9}$, $t_2 = \frac{5}{9}$ and $t_3 = 1$. Time and price increments are the ones that follow.

Absolute log-price increments	Time increments
$\left\vert \ln[S(t_1)] - \ln[S(t_0)] \right\vert = \left\vert \frac{2}{3} - 0 \right\vert = \frac{2}{3}$	$t_1 - t_0 = \frac{4}{9} - 0 = \frac{4}{9}$
$\left\vert \ln[S(t_2)] - \ln[S(t_1)] \right\vert = \left\vert \frac{1}{3} - \frac{2}{3} \right\vert = \frac{1}{3}$	$t_2 - t_1 = \frac{5}{9} - \frac{4}{9} = \frac{1}{9}$
$\left\vert \ln[S(t_3)] - \ln[S(t_2)] \right\vert = \left\vert 1 - \frac{1}{3} \right\vert = \frac{2}{3}$	$t_3 - t_2 = 1 - \frac{5}{9} = \frac{4}{9}$

Hence we find the general rule

$$\left\vert \Delta \ln[S(t)] \right\vert = \sqrt{\Delta t}\,,$$

which can be considered as a deterministic version of a SBM. As a matter of fact, from Definition 1.1.1, if $\{\ln[S(t)]\}_{t \in [0,1]}$ is a standard Brownian motion, we have

$$\ln[S(t)] - \ln[S(t - \Delta t)] \overset{d}{\sim} \mathcal{N}(0, \Delta t)\,.$$

If we set $\varepsilon \overset{d}{\sim} \mathcal{N}(0,1)$, we can write

$$\Delta \ln[S(t)] = \varepsilon \cdot \sqrt{\mathbb{V}\big(\Delta \ln[S(t)]\big)} = \varepsilon \cdot \sqrt{\Delta t}\,,$$

and hence in the 68% of the cases we have

$$\mu_{\Delta S}(t) - \sigma_{\Delta S}(t) \leq \Delta \ln[S(t)] \leq \mu_{\Delta S}(t) + \sigma_{\Delta S}(t)$$

$$-\sqrt{\Delta t} \leq \Delta \ln[S(t)] \leq \sqrt{\Delta t}$$

$$\left\vert \Delta \ln[S(t)] \right\vert \leq \sqrt{\Delta t}\,.$$

So using such a generator (usually called *square root rule*) we get a *mild* variability, very similar to the one of a coin tossing. More precisely, we create

the variations of one standard deviation exactly; SBM would allow even more restrained variations (the ones given by the $<$, so the prefix "pseudo"). Adherence to the assumptions behind this oversimplified model constitutes the same mistake made by following those hypotheses of classical finance.

A first and very important generalization of the previous model yields to ones that are non-Brownian but are called *unifractal*. They continue to require that the height of every segment of the generator be linked to its width, but by a function of an exponent H, that is

$$|\Delta \ln[S(t)]| = (\Delta t)^H$$

which is linked to the stochastic version

$$\Delta \ln[S(t)] = \varepsilon \cdot (\Delta t)^H \ ,$$

that is a standard FBM if $H \in (0,1)$ (but this restriction is not necessary to merely create cartoons that imitate financial prices).

A far more drastic generalization of the previous recursive models can be achieved introducing a "speed" in market activity. As a matter of fact, it can be assumed that a financial market speeds up and slows down, depending on the activity it exhibits. This variability is the essence of volatility. This is why models that allow for such form of variability change their prefix "uni" into "multi" before the word "fractal", hence leading to a *multifractal* model of asset return.

A precise definition of "activity" will be avoided, since it is unnecessary. The key-idea is that the market does not always follow the physical time, but instead a subjective time that passes slowly during some periods and fast during others.

The key-step is shown in Figure 3.4. The horizontal time axis is lengthened or shorten so that the prices of the generator are either stretched or squeezed. At the same time, the vertical price axis may remain untouched. As seen on the plane (CLOCK TIME , PRICE), the first part of the generator given by the unifractal model is progressively shortened, providing more space to lengthen the second part (note that the third piece of the generator remains unchanged). After making these progressive "adjustments", the generator become multifractal (passing from f_U to f_4). Moreover, seeing on the plane (CLOCK TIME , TRADING TIME), market activity speeds up in the interval of time represented by the first piece of the generator and slows in the interval that corresponds to the second piece.

The plane (TRADING TIME , PRICE) shows an oscillating generator in trading time. This is the pseudo-Brownian (unifractal) generator of Figure 3.3. The plane (CLOCK TIME , PRICE) shows four multifractal oscillating generators in clock time. Finally the last plane shows the relations between clock and trading time. If we consider the lowest (straight) line, we

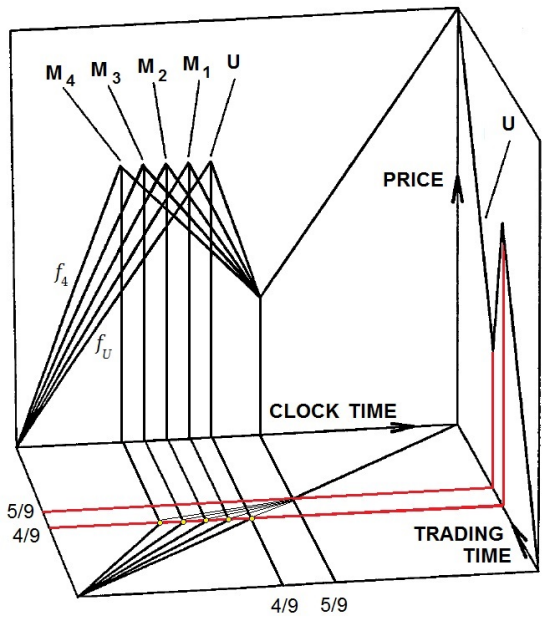


Figure 3.4: A *Multifractal* Model of Asset Return

do not see any differences between them, that is

$$\text{TRADING TIME} = \text{CLOCK TIME}\,.$$

However, introducing the multifractal generator, each becomes an increasing function of the other, that is that is

$$\text{TRADING TIME} = \theta\,(\text{CLOCK TIME})\ ,$$

being $\theta(\cdot)$ the increasing function. Moving a piece of the fractal generator to the left causes some amount of market activity to occur in a *shorter* time interval for the first piece of the generator.

The pattern of the previous figure surely does not exhaust all the possibilities offered by the theory of fractals and cartoons. The one of Figure 3.4 is obtained by a pseudo-Brownian generator, but more sophisticated plot can be easily drawn. Such an alteration to the generator can is able to produce a full simulation of price fluctuations over a given period, using the process of interpolations described earlier.

Furthermore, to make these models of volatile markets achieve the necessary realism, the three pieces of each generator have to be randomized. If we call the three pieces of them as G_1, G_2 and G_3 for the first, second and third part respectively, six permutations are allowed. These are

$$(G_1, G_2, G_3) \qquad (G_1, G_3, G_2)$$

$$(G_2, G_1, G_3) \qquad (G_2, G_3, G_1)$$
$$(G_3, G_1, G_2) \qquad (G_3, G_2, G_1)$$

where the particular case in Figure 3.3 is given by the first one (G_1, G_2, G_3). If each of these six elements are seen as the six sides of a die, to get a mild random model, before each interpolation, the die is thrown and the permutation that comes up is selected.

3.2 Binomial, multinomial and canonical measures

In order to understand the MMAR, which is based on multifractality, few preliminary notions such as *binomial* and *multinomial measures* are required. In the previous chapter we have introduced multifractal (measures) in a very theoretical fashion. In the following, we will introduce the simplest multifractal, the (*Bernoulli*) *binomial measure* on the compact interval $[0, 1] \subset \mathbb{R}$. Furthermore, we will construct further multifractal measures which derive from the former.

The recursive construction of the binomial measure involves an initiator and a generator. The initiator is the interval $[0, 1]$ itself on which a unit of (probability) mass is uniformly spread. This interval will recursively split into halves, leading to, at the k-th stage, dyadic intervals of length 2^{-k}. The generator consists in a single parameter $0 < u_0 < 1$ and $u_0 \neq \frac{1}{2}$, named *multiplier*, which at each stage is spread over the halves of every dyadic interval, with unequal deterministic proportions.

Let u_0 be a multiplier and be u_1 its ones' complement. At stage $k = 0$, we start the construction with the uniform probability measure on $[0, 1]$, that is[4]

$$f_0(t) = \begin{cases} \mathcal{M}_0([0,1]) = 1 & \text{if } t \in [0,1] \\ 0 & \text{if } t \notin [0,1] \end{cases}.$$

At the step $k = 1$,[5] the measure $\mathcal{M}_1$ uniformly spread mass equal to u_0 on the subinterval $\left[0, \frac{1}{2}\right]$ and mass equal to u_1 on $\left[\frac{1}{2}, 1\right]$, that is

$$f_1(t) = \begin{cases} \mathcal{M}_1\left(\left[0, \frac{1}{2}\right]\right) = u_0 & \text{if } t \in \left[0, \frac{1}{2}\right) \\ \mathcal{M}_1\left(\left[\frac{1}{2}, 1\right]\right) = u_1 & \text{if } t \in \left[\frac{1}{2}, 1\right] \end{cases}.$$

[4] If a random variable $X \overset{d}{\sim} \mathcal{U}(a, b)$ over its support $[a, b] \subset \mathbb{R}$, its density (which is the uniform probability measure) is given by

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{if } x \in [a,b] \\ 0 & \text{if } x \notin [a,b] \end{cases}.$$

[5] We omit the case $t \notin [0, 1]$, since we are introducing a measure on $[0, 1]$.

Here, is trivial to see that the mass is preserved. In fact

$$\mathcal{M}_1\left(\left[0,\frac{1}{2}\right]\right)+\mathcal{M}_1\left(\left[\frac{1}{2},1\right]\right)=u_0+u_1=1\,.$$

In step $k=2$, the set $\left[0,\frac{1}{2}\right]$ is split into two subintervals, $\left[0,\frac{1}{4}\right]$ and $\left[\frac{1}{4},\frac{1}{2}\right]$, which respectively receive a percentage u_0 and u_1 of the total mass $\mathcal{M}_1\left(\left[0,\frac{1}{2}\right]\right)$. Applying the same procedure to the dyadic set $\left[\frac{1}{2},1\right]$, we obtain

$$f_2(t)=\begin{cases}\mathcal{M}_2\left(\left[0,\frac{1}{4}\right]\right)=u_0\cdot u_0=u_0^2 & \text{if } t\in\left[0,\frac{1}{4}\right)\\ \mathcal{M}_2\left(\left[\frac{1}{4},\frac{1}{2}\right]\right)=u_0\cdot u_1=u_0\cdot u_1 & \text{if } t\in\left[\frac{1}{4},\frac{1}{2}\right)\\ \mathcal{M}_2\left(\left[\frac{1}{2},\frac{3}{4}\right]\right)=u_1\cdot u_0=u_0\cdot u_1 & \text{if } t\in\left[\frac{1}{2},\frac{3}{4}\right)\\ \mathcal{M}_2\left(\left[\frac{3}{4},1\right]\right)=u_1\cdot u_1=u_1^2 & \text{if } t\in\left[\frac{3}{4},1\right]\end{cases}.$$

As we can see the total mass is preserved since

$$\sum_{i=0}^{2^2-1}\mathcal{M}_2\left(\left[\frac{i}{2^2},\frac{i+1}{2^2}\right]\right)=u_0^2+2\cdot u_0\cdot u_1+u_1^2=(u_0+u_1)^2=1\,.$$

We want to ascertain that the procedure works even for $f_3(t)$.

$$f_3(t)=\begin{cases}\mathcal{M}_3\left(\left[0,\frac{1}{8}\right]\right)=u_0\cdot u_0\cdot u_0=u_0^3 & \text{if } t\in\left[0,\frac{1}{8}\right)\\ \mathcal{M}_3\left(\left[\frac{1}{8},\frac{1}{4}\right]\right)=u_0\cdot u_0\cdot u_1=u_0^2\cdot u_1 & \text{if } t\in\left[\frac{1}{8},\frac{1}{4}\right)\\ \mathcal{M}_3\left(\left[\frac{1}{4},\frac{3}{8}\right]\right)=u_0\cdot u_1\cdot u_0=u_0^2\cdot u_1 & \text{if } t\in\left[\frac{1}{4},\frac{3}{8}\right)\\ \mathcal{M}_3\left(\left[\frac{3}{8},\frac{1}{2}\right]\right)=u_0\cdot u_1\cdot u_1=u_0\cdot u_1^2 & \text{if } t\in\left[\frac{3}{8},\frac{1}{2}\right)\\ \mathcal{M}_3\left(\left[\frac{1}{2},\frac{5}{8}\right]\right)=u_1\cdot u_0\cdot u_0=u_0^2\cdot u_1 & \text{if } t\in\left[\frac{1}{2},\frac{5}{8}\right)\\ \mathcal{M}_3\left(\left[\frac{5}{8},\frac{3}{4}\right]\right)=u_1\cdot u_0\cdot u_1=u_0\cdot u_1^2 & \text{if } t\in\left[\frac{5}{8},\frac{3}{4}\right)\\ \mathcal{M}_3\left(\left[\frac{3}{4},\frac{7}{8}\right]\right)=u_1\cdot u_1\cdot u_0=u_0\cdot u_1^2 & \text{if } t\in\left[\frac{3}{4},\frac{7}{8}\right)\\ \mathcal{M}_3\left(\left[\frac{7}{8},1\right]\right)=u_1\cdot u_1\cdot u_1=u_1^3 & \text{if } t\in\left[\frac{7}{8},1\right]\end{cases},$$

and total mass still preserves

$$\sum_{i=0}^{2^3-1}\mathcal{M}_3\left(\left[\frac{i}{2^3},\frac{i+1}{2^3}\right]\right)=u_0^3+3\cdot u_0^2\cdot u_1+3\cdot u_0\cdot u_1^2+u_1^3=(u_0+u_1)^3=1\,.$$

This procedure can generate an infinite sequence of measures.

At step $k+1$, we assume that the measure $\mathcal{M}_k$ has been defined. To construct $\mathcal{M}_{k+1}$, consider a dyadic interval $\left[t,t+2^{-k}\right]$, where t is the *dyadic number* of the form:

$$t=(0.\eta_1\eta_2\ldots\eta_k)_2=\left(\sum_{i=1}^{k}\eta_i\cdot 2^{-k}\right)_{10}$$

for a finite k and $\eta_1, \eta_2, \ldots, \eta_k \in \{0, 1\}$ (hence we are using the counting base $b = 2$). Then we uniformly spread a fraction u_0 and u_1 of the mass $\mathcal{M}_k\left(\left[t, t + 2^{-k}\right]\right)$ on the subintervals $\left[t, t + 2^{-(k+1)}\right]$ and $\left[t + 2^{-(k+1)}, t + 2^{-k}\right]$. The repetition of this scheme to all the subintervals define the measure $\mathcal{M}_{k+1}$.

Let φ_0 and φ_1 denote the relative frequencies of 0's and 1's (that is $\varphi_1 = 1 - \varphi_0$) in the finite binary development $t = (0.\eta_1\eta_2 \ldots \eta_k)_2$. The so-called *pre-binomial measure* in the dyadic interval $\left[t, t + 2^{-k}\right]$, takes the value[6]

$$\mathcal{M}_k\left(\left[t, t + 2^{-k}\right]\right) = u_0^{k \cdot \varphi_0} \cdot u_1^{k \cdot \varphi_1} . \tag{3.1}$$

Because of the conservation of the mass at each stage, we can write

$$\sum_{i=0}^{2^k - 1} \mathcal{M}_k\left(\left[\frac{i}{2^k}, \frac{i+1}{2^k}\right]\right) = (u_0 + u_1)^k = 1 .$$

The iteration of the procedure generates an infinite sequence of random measure $\{\mathcal{M}_k\}$ that weakly converges to the *binomial measure* $\mathcal{M}$, that is

$$\mathcal{M}_k \xrightarrow{d} \mathcal{M} .$$

In fact, setting the length $\Delta t = \left(t + 2^{-k}\right) - t = 2^{-k}$, pre-binomial measure redistributes mass among the dyadic intervals, but the total mass is always preserved at each stage since $u_0 + u_1 = 1$.

Note that the binomial measure has important features common to many multifractal (measures): It is continuous but also a singular probability measure; it thus has no density[7]. Moreover it easy to prove it is a *self-similar* multifractal measure (in the sense of Definition 2.5.6). It will be shown in the next section.

This construction can receive several extensions. At each stage the interval can be split not in two but in $b > 2$ intervals of equal size. Subintervals, indexed from left to right by j $(0 \leq j \leq b - 1)$, receive fraction of the total

[6] Take as an instance $\mathcal{M}_3\left(\left[\frac{1}{8}, \frac{1}{4}\right]\right) = u_0^2 \cdot u_1$. Since $\frac{1}{8} = 0.125_{10}$ in base 10 (usual decimal representation), but in base 2 is equal to $t = 0.001_2$, we can see that the relative frequency of 0's is $\varphi_0 = \frac{2}{3}$ and that the relative frequency of 1's is $\varphi_1 = \frac{1}{3}$, using (3.1) we have

$$\mathcal{M}_3\left(\left[\frac{1}{8}, \frac{1}{4}\right]\right) = u_0^{3 \cdot \frac{2}{3}} \cdot u_1^{3 \cdot \frac{1}{3}} = u_0^2 \cdot u_1 ,$$

as expected.

[7] As a matter of fact, since $u_0, u_1, \varphi_0, \varphi_1 \in (0, 1)$, the limit

$$\lim_{k \to \infty} \mathcal{M}_k\left(\left[t, t + 2^{-k}\right]\right) = \lim_{k \to \infty} (u_0^{\varphi_0} \cdot u_1^{\varphi_1})^k = 0$$

being $0 < u_0^{\varphi_0} \cdot u_1^{\varphi_1} < 1$.

mass equal to $u_0, u_1, \ldots, u_{b-1}$. For conserving mass the multipliers have to be such that

$$\sum_{j=0}^{b-1} u_j = 1\,.$$

Thus the *multinomial measure* on the b-adic interval $\left[t, t+b^{-k}\right]$ follows the conservation rule

$$\sum_{i=0}^{b^k-1} \mathcal{M}_k\left(\left[\frac{i}{b^k}, \frac{i+1}{b^k}\right]\right) = \left(\sum_{j=0}^{b-1} u_j\right)^k = 1\,.$$

Here, the measure is computed as

$$\mathcal{M}_k\left(\Delta t\right) = \prod_{j=0}^{b-1} u_j^{k\cdot\varphi_j}\,. \tag{3.2}$$

where $\Delta t = b^{-k}$ and t is the b-adic number

$$t = (0.\eta_1\eta_2\ldots\eta_k)_b = \left(\sum_{i=1}^{k} \eta_i \cdot b^{-k}\right)_{10}$$

for a finite k and $\eta_1, \eta_2, \ldots, \eta_k \in \{0, 1, \ldots, b-1\}$, and φ_j are the relative frequencies of the digits of the representation in base b.

Moreover another discrete extension is easy to get, making the allocation of the mass random (the procedure is not so different by the one sketched at the end of the previous section with the metaphor of the die). Hence the multiplier of each subinterval is a sequence of *independent and identically distributed* (positive) random variables $\{U_j\}$. As for the previous cases, we need to assume that the mass is preserved *at each stage* of the construction[8], that is

$$\sum_{j=0}^{b-1} U_j = 1\,,$$

leading to the obvious fact $0 \leq U_j \leq 1$. Taking the expectation of the sum, we get an expression for the expected value of the single random variable,

[8] To create such a sequence of random variables, we can generate b random real numbers inside the interval $[0, 1]$. Then, we just have to put them in (increasing) order, and assign $U_j = t_j - t_{j-1}$ with $j = 0, 1, ..., b-1$ and $t_j \in [0, 1]$. By construction they sum up to one, being also independent.

that is[9]

$$\mathbb{E}(U) = \frac{1}{b}.$$

for all j. The resulting measure is called *microcanonical measure* (the reason of this name will be clear at a later stage). Given a date $t = (0.\eta_1\eta_2\ldots\eta_k)_b$ and a length $\Delta t = b^{-k}$, the measure of the b-adic cell $\left[t, t+b^{-k}\right]$ satisfies

$$\mathcal{M}_k(\Delta t) = U_{\eta_1} \cdot U_{\eta_1\eta_2} \cdot \ldots \cdot U_{\eta_1\eta_2\ldots\eta_k} \tag{3.3}$$

where $\eta_1\eta_2\ldots\eta_k$ is one of the element (it is a b-adic number) of the ordered selection with repetition[10] made with b digits. Because of its property, it follows

$$\mathcal{M}_k(\Delta t)^q = \left(U_{\eta_1} \cdot U_{\eta_1\eta_2} \cdot \ldots \cdot U_{\eta_1\eta_2\ldots\eta_k}\right)^q = U_{\eta_1}^q \cdot U_{\eta_1\eta_2}^q \cdot \ldots \cdot U_{\eta_1\eta_2\ldots\eta_k}^q$$

for all $q \geq 0$. Taking the expectation of this expression and considering the fact that the multipliers are independent (and hence their q-th power are independent as well), we get

$$\begin{aligned}\mathbb{E}\left[\mathcal{M}_k(\Delta t)^q\right] &= \mathbb{E}\left(U_{\eta_1}^q \cdot U_{\eta_1\eta_2}^q \cdot \ldots \cdot U_{\eta_1\eta_2\ldots\eta_k}^q\right) = \\ &= \mathbb{E}\left(U_{\eta_1}^q\right) \cdot \mathbb{E}\left(U_{\eta_1\eta_2}^q\right) \cdot \ldots \cdot \mathbb{E}\left(U_{\eta_1\eta_2\ldots\eta_k}^q\right).\end{aligned}$$

Since, in addiction, the random variables are identically distributed, the previous expression becomes the following scaling rule

$$\begin{aligned}\mathbb{E}\left[\mathcal{M}_k(\Delta t)^q\right] &= \mathbb{E}\left(U_{\eta_1}^q\right) \cdot \mathbb{E}\left(U_{\eta_1\eta_2}^q\right) \cdot \ldots \cdot \mathbb{E}\left(U_{\eta_1\eta_2\ldots\eta_k}^q\right) = \\ &= \underbrace{\mathbb{E}(U^q) \cdot \mathbb{E}(U^q) \cdot \ldots \cdot \mathbb{E}(U^q)}_{k \text{ times}} = \left[\mathbb{E}(U^q)\right]^k.\end{aligned}$$

Setting $\tau(q) \equiv -\log_b\left[\mathbb{E}(U^q)\right] - 1$ (which is named *scaling function*), the expression can be written as[11]

$$\mathbb{E}\left[\mathcal{M}_k(\Delta t)^q\right] = \Delta t^{\tau(q)+1} \tag{3.4}$$

---

[9] Since the unit mass is preserved, we have $\mathbb{E}\left(\sum_{j=0}^{b-1} U_j\right) = \mathbb{E}(1)$, and hence

$$\mathbb{E}\left(\sum_{j=0}^{b-1} U_j\right) = \mathbb{E}(U_0) + \mathbb{E}(U_1) + \ldots + \mathbb{E}(U_{b-1}) = \underbrace{\mathbb{E}(U) + \mathbb{E}(U) + \ldots + \mathbb{E}(U)}_{b \text{ times}} = 1.$$

[10] Hence the set $\{\eta_1, \eta_2, \ldots, \eta_k\}$ can originate $D'_{b,k} = b^k$ ordered selections with repetition.

[11] In fact, since $\Delta t = b^{-k}$, we have

$$\begin{aligned}\mathbb{E}(U^q)^k &= \left(\frac{1}{b}\right)^{\log_{1/b}\left[\mathbb{E}(U^q)^k\right]} = \left(\frac{1}{b}\right)^{-k\cdot\log_b\left[\mathbb{E}(U^q)\right]} = \left(b^{-1}\right)^{-k\cdot\log_b\left[\mathbb{E}(U^q)\right]} = \\ &= \left(b^{-k}\right)^{-1\cdot\log_b\left[\mathbb{E}(U^q)\right]} = \Delta t^{-\log_b\left[\mathbb{E}(U^q)\right]} = \Delta t^{-\log_b\left[\mathbb{E}(U^q)\right]-1+1} = \Delta t^{\tau(q)+1}.\end{aligned}$$

which is the typical behaviour of a multifractal measure[12].

Finally we are able to introduce the last multifractal measure, which will be the one involved in the Multifractal Model of Asset Returns. If, given a sequence of independently and identically distributed (positive) random variables, each iteration conserves probability mass only "on average" in the sense that

$$\mathbb{E}\left(\sum_{j=0}^{b-1} U_j\right) = 1\,,$$

we obtain a less restrictive case on multipliers, leading just to the positivity of them $U_j \geq 0$[13].

The corresponding measure is then called *canonical measure* and its total mass, denoted as Υ, is generally random. As a matter of fact, if we consider such a measure that has started with mass 1 in $[0,1]$, and has continued over infinite many stages, because of the lack of an exact conservation, the ultimate mass is not identical to 1, but is a random variable Υ.

An alternative way to see Υ is to add the masses in the b subcells of length $\frac{1}{b}$. In the first stage, the j-th interval is given the mass U_j. Ultimately, it contains the mass $\Upsilon_j \cdot U_j$, the quantities U_j and Υ_j being statistically independent. Hence, the total sum of the partial masses can be written as

$$\sum_{j=0}^{b-1} \Upsilon_j \cdot U_j \overset{d}{\sim} \Upsilon\,,$$

meaning they have the same distribution. Thus Υ is the fixed point of the operation of randomly weighted averaging using as weight the random quantities Υ_j (which are inevitably identically distributed).

Thus, given a time $t = 0.\eta_1\eta_2\ldots\eta_k$, at the k-th stage, the canonical measure of a b-adic interval surely generate the same effect as the microcanonical measure, that is

$$\mathcal{M}_k\left(\Delta t\right) = U_{\eta_1} \cdot U_{\eta_1\eta_2} \cdot \ldots \cdot U_{\eta_1\eta_2\ldots\eta_k}\,.$$

[12]See the analogy with the "definition" of multifractal that is, for an open ball with centre $\mathbf{x} \in \mathbb{R}^n$ and radius r,

$$\mathcal{M}[B(\mathbf{x}, r)] \approx r^{\alpha}\,.$$

Of course, since the microcanonical measure is a random measure, the relationship is valid for its expected value.

[13]Note that, since $\sum_{j=0}^{b-1} U_j = 1 \Rightarrow \mathbb{E}\left(\sum_{j=0}^{b-1} U_j\right) = 1$ (but the converse is not true), we have that the expectation of the random variables are all equal to

$$\mathbb{E}(U) = \frac{1}{b}$$

even in this case.

However, in contrast to the previous case, that is not all: Each stage is also subjected to the same process as has been for $[0,1]$. Therefore, the measure does *not* reduce to (3.3), but instead takes the form:

$$\mathcal{M}_k(\Delta t) = \Upsilon_{\eta_1\eta_2\ldots\eta_k} \cdot \Big(U_{\eta_1} \cdot U_{\eta_1\eta_2} \cdot \ldots \cdot U_{\eta_1\eta_2\ldots\eta_k} \Big) \tag{3.5}$$

Since $\Upsilon_{\eta_1\eta_2\ldots\eta_k} \overset{d}{\sim} \Upsilon$ and it is independent from U_j for all j, similarly as for the microcanonical measure, we can write

$$\mathbb{E}\Big[\mathcal{M}_k(\Delta t)^q\Big] = \mathbb{E}(\Upsilon^q) \cdot [\mathbb{E}(U^q)]^k \ ,$$

where the new moment prefactor $\mathbb{E}(\Upsilon^q)$ reflects high frequency effects due to scales lower than b^{-k}.
Setting $c(q) \equiv \mathbb{E}(\Upsilon^q)$, we finally get the sought expression

$$\mathbb{E}\Big[\mathcal{M}_k(\Delta t)^q\Big] = c(q) \cdot \Delta t^{\tau(q)+1} \, . \tag{3.6}$$

Far from being obvious, the scaling function "generated" by a canonical measure under the closed interval $[0,1]$ is always concave. For an extensive dissertation about that and other $\tau(q)$'s properties, see Appendix C.1. They will be crucial for a proper understanding of the MMAR.

3.3 Self-similar random measures

This section is devoted to the proof of the self-similarity of the canonical measure. In order to do this, we recall the definition of self-similar measure given at the end of the previous chapter. Here we provide a less general definition, valid for all the subset of the real line.

Definition 3.3.1 *Given a closed set $T \subset \mathbb{R}$ and a family of contracting similarities $\mathfrak{F} = \{\mathbf{F_1}, \mathbf{F_2}, ..., \mathbf{F_k}\}$ for which the the open set condition is fulfilled, a multifractal measure $\mathcal{M}$ is said to be a* self-similar measures *if*

$$\frac{\mathcal{M}[\mathbf{F_i}(T_1)]}{\mathcal{M}(T_1)} = \frac{\mathcal{M}[\mathbf{F_i}(T_2)]}{\mathcal{M}(T_2)} = \cdots = \frac{\mathcal{M}[\mathbf{F_i}(T_n)]}{\mathcal{M}(T_n)} \tag{3.7}$$

for every $i = 1, 2, \ldots, k$, where $T_1 \subseteq T_2 \subseteq \ldots \subseteq T_n$ is a non-decreasing sequence of compact sets contained in T.

Since the canonical measure involves random variables, it is a random measure. A random measure on the interval T is analogous to a random variable over the same support. It is a mapping defined on a probability space $(\Omega, \mathcal{F}, \mathbf{P})$, and valued in the class of all the measures on T. Moreover, given a fixed interval $T_i \subseteq T$, the mass $\mathcal{M}(T_i)$ is a random variable.

We now provide a good definition of self-similar random measure.

Definition 3.3.2 *Given a random measure $\mathcal{M}$ for which ratios of* (3.7) *are identically distributed, and such that, given a non-decreasing sequence of compact intervals $T_1 \subseteq T_2 \subseteq ... \subseteq T_n$ contained in T, the following ratios between random variable*

$$\frac{\mathcal{M}(T_1)}{\mathcal{M}(T_2)}, \frac{\mathcal{M}(T_2)}{\mathcal{M}(T_3)}, \ldots, \frac{\mathcal{M}(T_{n-1})}{\mathcal{M}(T_n)} \tag{3.8}$$

are statistically independent, the measure $\mathcal{M}$ is said to be a self-similar random measure.

Such definition means that successive iterations of the measure on the intervals are statistically independent. Given the interval $T \subseteq \mathbb{R}^+$, Definition 3.3.2 implies the existence of a positive stochastic process $\{F(\lambda)\}_{0<\lambda\leq 1}$ independent of $\mathcal{M}$ that satisfies

$$\mathcal{M}([0, \lambda \cdot t]) \overset{d}{\sim} F(\lambda) \cdot \mathcal{M}([0, t]) \tag{3.9}$$

whenever $t \in T$ (note that, by construction, $F(1) = 1$). Given two coefficients $a, b \in (0, 1]$, we can write

$$\frac{\mathcal{M}([0, ab \cdot t])}{\mathcal{M}([0, t])} = \frac{\mathcal{M}([0, ab \cdot t])}{\mathcal{M}([0, a \cdot t])} \cdot \frac{\mathcal{M}([0, a \cdot t])}{\mathcal{M}([0, t])} = \frac{\mathcal{M}([0, b \cdot t])}{\mathcal{M}([0, t])} \cdot \frac{\mathcal{M}([0, a \cdot t])}{\mathcal{M}([0, t])} .$$

The multiplicative measures of the previous sections are self-similar for such a set of value of t and λ. For instance, in the case of the binomial measure, the previous ratios are consistent if we choose t and λ as dyadic numbers (hence, in the case of the multinomial measures, we have to use b-adic numbers).

By virtue of (3.8), the two ratios on the right-hand side are statistically independent. Moreover, the process $\{F(\lambda)\}$ satisfies the following property

$$F(ab) \overset{d}{\sim} F(a) \cdot F(b) .$$

This statement implies that, given $q \geq 0$

$$\mathbb{E}\big[F(ab)^q\big] = \mathbb{E}\big[F(a)^q \cdot F(b)^q\big] = \mathbb{E}\big[F(a)^q\big] \cdot \mathbb{E}\big[F(b)^q\big] ,$$

being, by definition, $F(a)$ and $F(b)$ statistically independent.

We can now discuss the moments of the measure. When the moments are finite, the process $\{F(\lambda)\}$ satisfies the scaling relationship[14] :

$$\mathbb{E}\big[F(\lambda)^q\big] = \lambda^{\tau(q)+1} .$$

Letting $t = 1$, since

$$\mathcal{M}([0, \lambda])^q = F(\lambda)^q \cdot \mathcal{M}([0, 1])^q ,$$

[14] We have $\tau(q) = \log_\lambda\left(\mathbb{E}\left[F(a)^q\right]\right) + \log_\lambda\left(\mathbb{E}\left[F(b)^q\right]\right) - 1$, with $\lambda = a \cdot b$.

if we set $\mathcal{M}(\lambda) = \mathcal{M}([0,\lambda])$, therefore

$$\mathbb{E}\big[\mathcal{M}(\lambda)^q\big] = \mathbb{E}\big[F(\lambda)^q \cdot \mathcal{M}(1)^q\big] .$$

Moreover, since the stochastic process $\{F(\lambda)\}_{0<\lambda\leq 1}$ is independent of $\mathcal{M}$, we get

$$\mathbb{E}\big[\mathcal{M}(\lambda)^q\big] = \mathbb{E}\big[F(\lambda)^q\big] \cdot \mathbb{E}\big[\mathcal{M}(1)^q\big] = \mathbb{E}\big[\mathcal{M}(1)^q\big] \cdot \lambda^{\tau(q)+1} . \tag{3.10}$$

Since this last expression is the same as (3.6), we have proved that the canonical measure is a self-similar random measure. This results will give us a consistent justification in defining *multifractal processes* in terms of their moments.

Such approach to the multifractal measure over $[0,1]$ allows us to show further properties of the scaling function $\tau(q)$. Setting $q = 0$ we find

$$\mathbb{E}\big[\mathcal{M}(\lambda)^0\big] = \mathbb{E}\big[\mathcal{M}(1)^0\big] \cdot \lambda^{\tau(0)+1}$$

$$1 = \lambda^{\tau(q)+1} .$$

which imposes $\tau(0) = -1$. Moreover, if we consider the interval $[0,1]$, the total mass is represented, by definition, by $\mathcal{M}(1)$. Partitioning the interval in n subintervals of equal length, the total mass must be, on average, equally spread over the n cells. Because the total mass does not vanish but is conserved, we have necessarily

$$\frac{\mathbb{E}\big[\mathcal{M}(1)\big]}{n} = \mathbb{E}\left[\mathcal{M}\left(\frac{1}{n}\right)\right] .$$

Setting $q = 1$, we find

$$\mathbb{E}\big[\mathcal{M}(\lambda)\big] = \mathbb{E}\big[\mathcal{M}(1)\big] \cdot \lambda^{\tau(1)+1} = n \cdot \mathbb{E}\left[\mathcal{M}\left(\frac{1}{n}\right)\right] \cdot \lambda^{\tau(1)+1} .$$

Setting $\lambda = \frac{1}{n}$, we have

$$\mathbb{E}\left[\mathcal{M}\left(\frac{1}{n}\right)\right] = n \cdot \mathbb{E}\left[\mathcal{M}\left(\frac{1}{n}\right)\right] \cdot \left(\frac{1}{n}\right)^{\tau(1)+1} ,$$

which is satisfied only if $\tau(1) = 0$.

We have to remark that the this second property $\tau(1) = 0$ is valid *only on average* for the microcanonical and canonical measure, being the random variable identically distributed. On the other hand, the intercept of $\tau(q)$ is always at -1, by construction.

3.4 Multifractal processes and local Hölder exponents

Now the concept of multifractality can be extended to stochastic process. Because of the previous introduction on multifractal random measures, we find convenient defining *multifractal processes* in terms of their moments. Nevertheless, we have to remark that dealing with measures rather than stochastic process may be similar, but it is not the same thing. In the following, we will discuss about those discrepancies.

Definition 3.4.1 *Given a filtered probability space* $(\Omega, \mathcal{F}, \mathbf{P})$*, every real stochastic process* $\{Y(t)\}_{t\in[0,+\infty)}$ *defined on it is called a* multifractal process*, if it has stationary increments and satisfies the following properties:*

(a) $Y(0) = 0$ *almost surely;*

(b) *The expectation of its absolute increments raised to the q-th power are such that*

$$\mathbb{E}\big[|Y(t) - Y(s)|^q\big] = c(q) \cdot |t - s|^{\tau(q)+1}$$

where $q \in Q \subseteq \mathbb{R}$ *and* $c, \tau : Q \to \mathbb{R}$.

Thank to the definition, setting $s = 0$, we get

$$\mathbb{E}\big[|Y(t)|^q\big] = c(q) \cdot t^{\tau(q)+1}\,. \tag{3.11}$$

Such a definition of multifractal process extend the one of self-affine process (whose standard and fractional Browninan motions belong to). As a matter of fact, since a self-affine process is a process such that, given $a > 0$

$$\{Y(a \cdot t)\}_{t\geq 0} \overset{d}{\sim} \left\{a^H \cdot Y(t)\right\}_{t\geq 0}$$

with $0 < H < 1$, it also satisfies[15]

$$\{Y(t)\}_{t\geq 0} \overset{d}{\sim} \left\{t^H \cdot Y(1)\right\}_{t\geq 0}\,.$$

Hence, it ensues

$$\{|Y(t)|\}_{t\geq 0} \overset{d}{\sim} \left\{t^H \cdot |Y(1)|\right\}_{t\geq 0}$$

[15] This is due to setting $a \cdot t = 1$, and applying it to the definition, that is

$$\{Y(1)\}_{t\geq 0} \overset{d}{\sim} \left\{\left(\frac{1}{t}\right)^H \cdot Y(t)\right\}_{t\geq 0},$$

$$\left\{t^H \cdot Y(1)\right\}_{t\geq 0} \overset{d}{\sim} \{Y(t)\}_{t\geq 0}\,.$$

and

$$\{|Y(t)|^q\}_{t\geq 0} \overset{d}{\sim} \left\{t^{Hq}\cdot|Y(1)|^q\right\}_{t\geq 0}\,.$$

Taking their expectation, we get

$$\mathbb{E}\big[|Y(t)|^q\big] = t^{Hq}\cdot\mathbb{E}\big[|Y(1)|^q\big]$$

that is a multifractal process with $\tau(q) = Hq - 1$ and $c(q) = \mathbb{E}\big[|Y(1)|^q\big]$. In the special case of self-affine processes, the scaling function $\tau(q)$ is linear and fully determined by the unique Hurst exponent H. Multifractal processes with linear $\tau(q)$ are called *unifractal* (recall what stated in Section 3.1). In this section, however, we will focus on multifractal processes with *non-linear* (concave) functions $\tau(q)$.

Here we explicit the first difference between a multifractal measure on a bounded interval (like $[0,1]$ in the previous sections) and a multifractal process on the real line.

Let q_1, q_2 be to exponent, and a_1, a_2 two positive number such that $a_1 + a_2 = 1$. By virtue of Hölder's inequality, we find[16]

$$\mathbb{E}\big[|Y(t)|^q\big] \leq \mathbb{E}\big[|Y(t)|^{q_1}\big]^{a_1}\cdot\mathbb{E}\big[|Y(t)|^{q_2}\big]^{a_2}\,,$$

where $q = a_1\cdot q_1 + a_2\cdot q_2$. Taking logarithms and using (3.11), we get[17]

$$\ln\left[c(q)\right] + \tau(q)\cdot\ln(t) \leq a_1\cdot\ln\left[c(q_1)\right] + a_2\cdot\ln\left[c(q_2)\right] + \Big[a_1\cdot\tau(q_1) + a_2\cdot\tau(q_2)\Big]\cdot\ln(t)\,.$$

If $t\in(0,1)$ (that is a bounded interval), dividing by $\ln(t) < 0$, we have

$$\frac{\ln\left[c(q)\right]}{\ln(t)} + \tau(q) \geq a_1\cdot\frac{\ln\left[c(q_1)\right]}{\ln(t)} + a_2\cdot\frac{\ln\left[c(q_2)\right]}{\ln(t)} + a_1\cdot\tau(q_1) + a_2\cdot\tau(q_2)\,.$$

Letting $t\to 0^+$, we find

$$\tau(q) \geq a_1\cdot\tau(q_1) + a_2\cdot\tau(q_2)\,,$$

[16] For the probability space $(\Omega,\mathcal{F},\mathbf{P})$, given two random variables A and B on Ω, Hölder's inequality states

$$\mathbb{E}\big(|A|\cdot|B|\big) \leq \big[\mathbb{E}\big(|A|^p\big)\big]^{\frac{1}{p}}\cdot\big[\mathbb{E}\big(|B|^s\big)\big]^{\frac{1}{s}}\,,$$

where $p\geq 1$ and $s < +\infty$, with $\frac{1}{p}+\frac{1}{s} = 1$. Setting $|A| = |Y(t)|^{a_1\cdot q_1}$, $|B| = |Y(t)|^{a_2\cdot q_2}$, and $\frac{1}{p} = a_1$, $\frac{1}{s} = a_2$, the inequalities above follows.

[17] Developing all the steps

$$\ln\Big[c(q)\cdot t^{\tau(q)+1}\Big] \leq \ln\Big\{\Big[c(q_1)\cdot t^{\tau(q_1)+1}\Big]^{a_1}\Big\} + \ln\Big\{\Big[c(q_2)\cdot t^{\tau(q_2)+1}\Big]^{a_2}\Big\}\,,$$

$$\ln\left[c(q)\right] + \left[\tau(q)+1\right]\cdot\ln(t) \leq a_1\cdot\Big\{\ln\left[c(q_1)\right] + \left[\tau(q_1)+1\right]\cdot\ln(t)\Big\} + a_2\cdot\Big\{\ln\Big[c(q_2)\Big] + \left[\tau(q_2)+1\right]\cdot\ln(t)\Big\}\,,$$

$$\ln\left[c(q)\right] + \tau(q)\cdot\ln(t) \leq a_1\cdot\ln\left[c(q_1)\right] + a_2\cdot\ln\left[c(q_2)\right] + \Big[a_1\cdot\tau(q_1) + a_2\cdot\tau(q_2)\Big]\cdot\ln(t) + (a_1+a_2-1)\cdot\ln(t)\,,$$

$$\ln\left[c(q)\right] + \tau(q)\cdot\ln(t) \leq a_1\cdot\ln\left[c(q_1)\right] + a_2\cdot\ln\left[c(q_2)\right] + \Big[a_1\cdot\tau(q_1) + a_2\cdot\tau(q_2)\Big]\cdot\ln(t)\,.$$

which confirms the concavity of $\tau(\cdot)$, rigorously proven in Appendix C.1.

However, if we we try to extend this procedure to the whole real line (as for a multifractal stochastic process defined as in Definition 3.4.1), the situation changes notably. In fact, if we divide the previous inequality by $\ln(t)$ with $t \in (1, +\infty)$, and then let $t \to +\infty$, we get the reverse inequality of (3.4)

$$\tau(q) \leq a_1 \cdot \tau(q_1) + a_2 \cdot \tau(q_2)\,.$$

So, for a stochastic process defined on $[0, +\infty)$, the function $\tau(q)$, being unique, is necessarily linear. Furthermore, this "imposes" the process to be unifractal.

An alternative explanation might be the following: If we consider a stochastic process (or measure) on a bounded interval, it is possible to find and distinguish (a continuum of) different scaling exponents, since there is "not so much" information involved. Differently, when $t \to +\infty$, these little different behaviours tends to disappear, standing out only the *most likely* exponent, which becomes the "driver" for the whole process.

Thus multifractality can only be reached for *bounded* time intervals or, alternatively speaking, processes defined on unbounded intervals are multifractal only on *bounded* ranges of time. This technical hitch has little consequence for financial modelling.

That phenomenon is what physicists call *crossover*[18]. Because of $\tau(q)$'s properties, which are related to the fact that the measure is canonical, the crossover displays a change in the distribution of the process, passing from a scaling Paretian behaviour (heavy tails) to a Gaussian one (thin tails). As t goes to infinity, the "rare" events end up disappearing. The same effect has been recognized and motivated also in more recent articles, such as [1], [2], and [3].

The table summarizes that circumstance.

Set of times	Process $Y(t)$	Scaling function $\tau(q)$
$t \in (0, 1)$	multifractal	concave
$t \in (1, +\infty)$	unifractal	linear

Moreover, the fact that in the uniscaling case (that is $\tau(q) = Hq - 1$) the crucial exponent is (the only one) Hurst exponent, it should be expected that the (continuum of) exponent of the multiscaling case might be related to the first derivative $\tau'(q)$ of the scaling function. This point well be analyzed hereinafter, though the multifractal spectrum and the Legendre transform.

[18] *Crossover* is a general term describing a situation when a system can go from being in one phase to being in another phase as a certain parameter is changed

Eventually, in order to study further properties of $\tau(q)$ and to be allowed to use some results of multifractal formalism introduced in the previous chapter, we need the definition of *local Hölder exponent* of a function. This is connected with Hölder function (for the definition see Section 2.3), but it characterized the smoothness of a function at a *given point*.

Definition 3.4.2 *Let $g : \mathbb{R} \supseteq A \to \mathbb{R}$ be a function defined on a neighborhood of a given point t_0. The number*[19]

$$\alpha(t_0) := \sup\left\{\beta > 0 : |g(t_0 + h) - g(t_0)| = O\left(|h|^{\beta}\right)\right\} \tag{3.12}$$

as $h \to 0$, is called the local Hölder exponent *of g at t_0.*

We note that $\alpha(t_0)$ is non-negative if and only if the function g is bounded around t_0. In the following, we will consider only this case[20]. Local Hölder exponent is a notion that can be well applied to functions and measures, deterministic or stochastic, with some adjustment. In the previous chapter we have seen that SBM and FBM have local Hölder exponent equal to $\frac{1}{2}$ and H respectively. As a matter of fact, continuous stochastic process are characterized by a *unique* Hölder exponent. Differently, in the following we will examine the case when a continuum of Hölder exponent is allowed.

From Definition 3.4.2 it easy to see that, for a continuous function $g(t)$, the Hölder exponent at time $t = t_0$ can be computed as[21]

$$\alpha(t_0) = \limsup_{h \to 0} \frac{\log|g(t_0 + h) - g(t_0)|}{\log|h|}.$$

This expression suggests a method for estimating the probability that a point, randomly chosen on the interval $[0, 1]$, will have a given Hölder exponent. An essential simplification for both, analytical and empirical study, is to replace the previous continuous limit by a discrete one.

[19] Given $g(x) \neq 0$, we say $f(x) = O(g(x))$ in a neighborhood x_0 if

$$\limsup_{x \to x_0} \left|\frac{f(x)}{g(x)}\right| < +\infty.$$

[20] This is an application of coarse multifractal analysis. However, in Appendix ?? we will show that using fine multifractal analysis, though the fine multifractal spectrum, negative values of α can be reached.

[21] In fact, since $|g(t_0 + h) - g(t_0)| = O\left(|h|^{\beta}\right)$, it follows

$$\log|g(t_0 + h) - g(t_0)| = O\left(\log\left(|h|^{\beta}\right)\right)$$

$$\log|g(t_0 + h) - g(t_0)| = O(\beta \cdot \log|h|)$$

$$\log|g(t_0 + h) - g(t_0)| = \beta \cdot O(\log|h|).$$

Given the set $[0,1] \subset \mathbb{R}$, we iteratively subdivide the interval in b^k equal size pieces, k denoting the stage in the sequence of subdivision. At each stage, at point $t = t_i \in [0,1]$, we compute the finite quantities $\left|g\left(t_i + b^{-k}\right) - g\left(t_i\right)\right|$ for each b^k subdivision. We define *coarse Hölder exponent*, the following quantity

$$\alpha_k\left(t_i\right) := \frac{\log\left|g\left(t_i + b^{-k}\right) - g\left(t_i\right)\right|}{\log\left(b^{-k}\right)},$$

and hence the Hölder exponent at point t_i is given by

$$\alpha\left(t_i\right) := \liminf_{k\to\infty} \frac{\log\left|g\left(t_i + b^{-k}\right) - g\left(t_i\right)\right|}{\log\left(b^{-k}\right)}.$$

From the definition we see that, varying the length of the interval $\Delta t = b^{-k}$, we can find different values of $\alpha\left(t_i\right)$. Partitioning the range of them into small non-overlapping intervals $(\alpha_j, \alpha_j + \varepsilon]$, and denoted by $N_k(\alpha_j, \varepsilon)$ the *number* of coarse Hölder exponents contained in each interval $(\alpha_j, \alpha_j + \varepsilon]$, as $k \to \infty$, the ratio $\frac{N_k(\alpha,\varepsilon)}{b^k}$ converges to the probability that a randomly selected point t_i has local Hölder exponent equal to $\alpha\left(t_i\right)$.

Even if this approach of representing the distribution of different Hölder exponents is correct, it will fail in a multifractal contest, since it is not *able* to distinguish between multifractal and unifractal processes. As a matter of fact, multifractals, allowing different values of α, typically have that a single Hölder exponent which predominates (called α^*), in the sense that the set of points $T_{\alpha^*} \subset [0,1]$ with exponent α^* "usurps" all of the *Lebesgue measure*. Differently, most of a *multifractal measure* concentrates on a set of instants with Hölder exponent differen from α^*. In order to distinguish between the two cases we need the following function.

Definition 3.4.3 *Given a function $g : [0,1] \supseteq A \to \mathbb{R}$, using the same iterative procedure used to compute coarse Hölder exponents and using the same notation, we define*

$$f(\alpha) := \lim_{\varepsilon\to 0} \liminf_{k\to\infty} \frac{\log\left[N_k(\alpha,\varepsilon)\right]}{\log\left(\frac{1}{b^{-k}}\right)} = \lim_{\varepsilon\to 0} \liminf_{k\to\infty} \frac{\log\left[N_k(\alpha,\varepsilon)\right]}{\log\left(b^k\right)}. \qquad (3.13)$$

If the limit exists and is positive (on a support larger than a single point), we call the function g a multifractal[22].

Of course the same definition can be applied substituting g with a random function, such as one of the (random) measures $\mathcal{M}$ defined in Section 3.2.

This approach tells us that there exist different fractal local scaling behaviors as measured by different local coarse Hölder exponents, being so not

[22] This "new" definition of multifractal is connected to the one of (*lower*) *coarse multifractal spectrum* ($f(\alpha) \equiv \underline{f}_C(\alpha)$). To see this, compare with Definition 2.5.1

uniform in general. In other words, $\alpha(t)$ is typically not constant in t but assume a whole range of values, thus imprinting a rich structure on the object of interest. This structure can be characterized either in geometrical terms making use of the concept of dimension, or in statistical terms based on sample moments. A tight connection between these two descriptions will emerge from the multifractal formalism.

In fact, many authors interpreted $f(\alpha)$ as the Hausdorff dimension of the set of points having local Hölder exponent equal to α. For any $\alpha \geq 0$, we can define a set T_α of instants with Hölder exponent α. As any subset of the real line, T_α has Hausdorff dimension $0 \leq \dim_{\mathcal{H}}(T_\alpha) \leq 1$[23]. Since we have shown that the microcanonical and canonical measure are self-similar random measures, from Section 2.5, applying (2.43), we know that

$$f(\alpha) = \dim_{\mathcal{H}}(T_\alpha) \ . \tag{3.14}$$

Thanks to the same result we know that $f(\alpha)$ is also the Legendre transform

$$f(\alpha) = \tau^\star(\alpha) = \inf_{q \in \mathbb{R}} \{q \cdot \alpha - \tau(q)\}$$

where $\tau(q) \equiv -\log_b [\mathbb{E}(U^q)] - 1$ is the well-known scaling function[24] and $\tau^\star(\alpha)$ denotes its Legendre transform.

In order to understand this mathematical tool properly, see Appendix C.2 where general properties of it, and few related to our subject of interest, are discussed.

For our field of interest, closed-form expressions for $f(\alpha)$ depends on the densities of the random variables $\{U_j\}$ involved in the multifractal measure $\mathcal{M}$. However, in order to find them, an important theorem of *Large Deviation Theory* is required. To examine in depth that properties, see Appendix C.3.

3.5 The MMAR

The Multifractal Model of Asset Returns appeared for the first time in the three paper series [7], [8], and [9], introducing the concept of multifractality to economics. This model attempts to describe price changes, accounting for several features of financial data: Long memory, fat tails and scale invariance. The authors especially criticized the GARCH-type representations, the latter assuming that the conditional distribution of the return (in respect to the information available until today) has a finite, time-varying second moment. Given the time series of the returns $\{x_i\}$[25] with $i = 1, 2, ..., t-1$, if we set

[23] Note the analogy with definition of fine Hausdorff multifractal spectrum of Definition 2.5.4.

[24] As a matter of fact, it can be easily shown that $\tau(q) = -\beta(q)$ where $\beta(q)$ is defined as in (2.37).

[25] Hence $\{x_i\}$ is the empirical realization of the discrete stochastic process $\{X_i\}$, where X_i is the random log-return at time i.

$h_t := \mathbb{V}(X_t \,|\, \mathcal{F}_{t-1})$, it follows

$$h_t = f(h_{t-j}, \varepsilon_{t-j})$$

where $\varepsilon_t \,|\, \mathcal{F}_{t-1} \overset{d}{\sim} \mathcal{N}(0,1)$ and j an integer number. Moreover, $f(\cdot)$ is assumed to be a bounded ARMA process, which merely is an affine function of its arguments. Note that such a model directly addresses volatility clustering in the data, creating heavy tails. However, neither long-range dependence nor scale invariance can be described with it. Furthermore, being scale invariance the equivalence between representations of the model at different time-scales, the authors remark that the absence of an invariance under scaling implies that, in empirical works, the researcher adds an *additional restriction* to the model when choosing the time-scale of the data.

In this way, temporal heterogeneity is introduced by time-varying conditional second moments in a discrete time framework. Conversely, multifractality introduces a new source of heterogeneity through time-varying local regularity in the price path. The concept of local H older exponent is able to describe local regularity. Multifractal processes bridge the gap between locally Gaussian diffusions and jump-diffusions by allowing a multiplicity of Hölder exponents.

As a matter of fact, the MMAR generate heavy tails and a divergent variance directly in the directing process of log-returns. Prices of financial asset are seen as a *multiscaling process* with long memory and heavy tails. Persistence in volatility is given by the use of a *random trading time*, generated as the cumulative distribution function of a random multifractal measure (which, we remark, can be seen as a random variable).

Additionally, the authors tested and simulated the model generating very realistic sample paths. It might open the door to new theoretical and applied approaches to asset pricing and risk valuation.

As we have already sketched, trading time plays a notable robe in transmitting multifractality to financial records. We do need a preliminary definition regarding a particular class of stochastic processes.

Definition 3.5.1 *Let $\{Z(t)\}_{t\in[0,+\infty)}$ be a stochastic process and $\theta(t)$ an increasing function of the time t. The process*

$$X(t) := Z[\theta(t)] \tag{3.15}$$

is called a compound process.

Since t denotes the clock physical time, the function $\theta(t)$ reproduces the so-called trading time, as in Section 3.1. We are now able to state the theoretical assumption of the MMAR.

Let $X(t)$ be the stochastic process describing rates of log-return, that is

$$X(t) := \ln[S(t)] - \ln[S(0)]$$

where $\{S(t)\}_{t\in[0,T]}$ is the stochastic process leading the price of the financial asset, where $T > 0$ is the final time[26]. The MMAR bases on the following three hypotheses:

H1 $\{X(t)\}_{t\in[0,T]}$ is a compound process

$$X(t) = B_H[\theta(t)]$$

where $B_H(t)$ is a fractional Brownian motion with Hurst exponent H, and $\theta(t)$ is a stochastic trading time;

H2 The trading time $\theta(t)$ is the cumulative distribution function of a multifractal measure defined on $[0,T]$[27]. Thus, $\theta(t)$ is a multifractal process with continuous, non-decreasing paths, and stationary increments;

H3 $\{B_H(t)\}_{t\in[0,T]}$ and $\{\theta(t)\}_{t\in[0,T]}$ are independent.

Hence trading time plays a crucial role in the MMAR. Moreover, compounding allows direct modelling of a process' variability without affecting the direction of the increments or their correlation[28].

We first note that $\theta(0)$ must be null due to the definition of $X(t)$, which imposes $X(0) = 0$. Moreover **H2** imposes that $\theta(t)$ be the cumulative distribution function of a self-similar random measure, such as the microcanonical or the canonical.

It is quite natural to expect that trading time $\theta(t)$ "transfers" multifractality to $X(t)$, and that scaling functions $\tau_\theta(q)$ and $\tau_X(q)$ may be related. The following theorem expresses this intuition.

Theorem 3.5.2 *Under hypotheses* **H1**, **H2**, **H3**, *the process $X(t)$ is multifractal, with stationary increments and scaling function*

$$\tau_X(q) = \tau_\theta(Hq)\,. \tag{3.16}$$

Proof: Since

$$X(t) = B_H[\theta(t)]\,,$$

[26] The reason of the introduction of a final time T, rather than taking into account a period $[0,+\infty)$, is due to multifractality which may exist only on bounded intervals. See Section 3.4.

[27] All the measure introduced in Section 3.2 are defined on $[0,1]$. With little effort, they can be extended to the set of times $[0,T]$, setting $\Delta t = b^{-k}\cdot T$, being t the usual b-adic number.

[28] There are several authors that prefer using *trading volumes* instead of trading time as deforming process.

using the law of total expectation, $\mathbb{E}\big[|X(t)|^q\big]$ can be written as

$$\mathbb{E}\big[|X(t)|^q\big] = \mathbb{E}\Big\{\mathbb{E}\big[|X(t)|^q \,|\, \theta(t) = u\big]\Big\}.$$

Due to independence between trading time $\{\theta(t)\}_{t\in[0,T]}$ and the fractional Brownian motion $\{B_H(t)\}_{t\in[0,T]}$, applying (1.17) we have

$$\mathbb{E}\big[|X(t)|^q \,|\, \theta(t) = u\big] = \mathbb{E}\big[|B_H(u)|^q \,|\, \theta(t) = u\big] = \mathbb{E}\big[|B_H(1)|^q\big] \cdot \theta(t)^{Hq}.$$

Thus,

$$\mathbb{E}\big[|X(t)|^q\big] = \mathbb{E}\left\{\mathbb{E}\big[|B_H(1)|^q\big] \cdot \theta(t)^{Hq}\right\} = \mathbb{E}\big[|B_H(1)|^q\big] \cdot \mathbb{E}\left[\theta(t)^{Hq}\right].$$

Since $\theta(t)$ satisfies the scaling relation

$$\mathbb{E}\big[|\theta(t)|^q\big] = c_\theta(q) \cdot t^{\tau_\theta(q)+1},$$

we have

$$\mathbb{E}\big[|X(t)|^q\big] = \mathbb{E}\big[|B_H(1)|^q\big] \cdot c_\theta(Hq) \cdot t^{\tau_\theta(Hq)+1}.$$

Setting $c_X(q) = \mathbb{E}\big[|B_H(1)|^q\big] \cdot c_\theta(Hq)$ and $\tau_X(q) = \tau_\theta(Hq)$, we find

$$\mathbb{E}\big[|X(t)|^q\big] = c_X(q) \cdot t^{\tau_X(q)+1}, \tag{3.17}$$

which is the thesis.

□

Hence, choosing a fractional Brownian motion as compounder and a multifractal measure to deform time, we are able to "spread" multiscaling to the process of the asset return $X(t)$. This is one of the most important property of the MMAR. In the following we analyze the other features of such a process.

Firstly, by (3.17) we see that, if $\mathbb{E}\big[|X(t)|^q\big]$ is finite for some t, then it is finite for all t. Moreover, since $\mathbb{E}\big[|X(t)|^q\big]$ depends on $\mathbb{E}\left[\theta(t)^{Hq}\right]$, it is finite if and if only the process $\theta(t)$ has finite moment of order Hq. Hence, the trading time *controls* the moment of the return $X(t)$.

Due to **H2**, the trading time is generated by a self-similar random measure. It is crucial to highlight that, whether the measure is microcanonical or canonical, the moments possess very different features. As a matter of fact, microcanonical measures have a fixed mass on the interval $[0,T]$ on which are defined. So $\theta(t)$ is bounded, and the compound process $X(t)$ has finite moments of all order. As Mandelbrot said in [7] and [27], it generates *mild* randomness, with relatively thin tails.

On the other hand, being canonical measure depending on the random variable Υ, it permits the model to have divergent moments. For an in-depth understanding see Appendix C.1. The corresponding process $X(t)$ will be

then *wild*. Overall, the MMAR has enough flexibility to allow for a wide variety of tail behaviour, both thin and fat.

We are now interested in martingale property of the MMAR. The following result follows.

Theorem 3.5.3 *If* $\{B_H(t)\}_{t\in[0,T]}$ *is a standard Brownian motion* $\left(H=\frac{1}{2}\right)$, *the following properties hold true:*

1. *If* $\mathbb{E}\left[\theta(t)^{\frac{1}{2}}\right]$ *is finite, then* $\{X(t)\}_{t\in[0,T]}$ *is a martingale with respect to its natural filtration;*

2. *If* $\mathbb{E}[\theta(t)]$ *is finite, the increments of* $\{X(t)\}_{t\in[0,T]}$ *are uncorrelated.*

Proof:

1. Let $\mathcal{F}_t^X$ and $\mathcal{F}_t^{(X,\theta)}$ denote respectively the natural filtration of $\{X(t)\}$ and $\{X(t),\theta(t)\}$. By the law of total expectation, given $s\leq t$ we have

$$\mathbb{E}\Big[X(t)\,|\,\mathcal{F}_s^X\Big]=\mathbb{E}\Big\{\mathbb{E}\left[B_H[\theta(t)]\,|\,\mathcal{F}_s^{(X,\theta)},\theta(t)=u\right]|\,\mathcal{F}_s^X\Big\}\,.$$

   For any $s\leq t\leq u$, the independence between $\{B_H(t)\}$ and $\{\theta(t)\}$ implies that, if $H=\frac{1}{2}$:

$$\mathbb{E}\left[B_H[\theta(t)]\,|\,\mathcal{F}_s^{(X,\theta)},\theta(t)=u\right]=\mathbb{E}\left[B_H(u)\,|\,\mathcal{F}_s^{(X,\theta)}\right]=B_H[\theta(s)]\,,$$

   being $\left\{B_{\frac{1}{2}}(t)\right\}$ (SBM) a martingale by Theorem 1.1.11. Hence we find that[29]

$$\mathbb{E}\Big[X(t)\,|\,\mathcal{F}_s^X\Big]=X(s)\,.$$

2. If $\mathbb{E}\big[X(t)^2\big]<+\infty$, it is a square integrable martingale, hence it has also uncorrelated increments[30]. Moreover, since we are dealing with Brownian motion, Theorem 1.1.3 (2), has already proved that the incremental process has uncorrelated increments (rather, it has independent increments due to Gaussianity).

---

[29] Since we are computing $\mathbb{E}\Big[X(t)\,|\,\mathcal{F}_s^X\Big]$, we implicitly are setting $q=1$. For this reason, we know that the trading time (and hence the Brownian motion) has finite moments only if $\mathbb{E}\left[\theta(t)^{Hq}\right]=\mathbb{E}\left[\theta(t)^{\frac{1}{2}}\right]<+\infty$, having set $H=\frac{1}{2}$. The finiteness of the moments is required for being a martingale. This is the reason for the hypothesis in the theorem's statement.

[30] Requiring $\mathbb{E}\big[X(t)^2\big]<+\infty$ is the same as stating $\mathbb{E}[\theta(t)]<+\infty$, since here $q=2$ and $H=\frac{1}{2}$.

□

This result is based on the martingale property of the Brownian motion. As it may be easily imagined, it does not extend when $H \neq \frac{1}{2}$. In this case, the following holds.

Theorem 3.5.4 *If $\{B_H(t)\}_{t\in[0,T]}$ is a fracional Brownian motion, and $\mathbb{E}\big[\theta(t)^{2H}\big]$ is finite, the autocovariance function at points t_1, t_2 of the incremental process $V(t) := X(t+\Delta t) - X(t)$, with $t \geq \Delta t$, is equal to*

$$c_\theta(2H) \cdot \frac{\sigma^2}{2} \cdot \Big(|t_1 - t_2 + \Delta t|^{\tau_\theta(2H)+1} + |t_1 - t_2 - \Delta t|^{\tau_\theta(2H)+1} - 2 \cdot |t_1 - t_2|^{\tau_\theta(2H)+1}\Big),$$

where $\sigma^2 = \mathbb{E}\left[B_H(1)^2\right]$.

Proof: Let we consider two different times t_1 and t_2. The autocovariance function of the incremental process is given by

$$\mathcal{C}_V(t_1,t_2) = \mathbb{E}\left[V(t_1)\cdot V(t_2)\right] = \mathbb{E}\left\{[X(t_1+\Delta t) - X(t_1)]\cdot[X(t_2+\Delta t) - X(t_2)]\right\}.$$

Since $X(t) = B_H[\theta(t)]$, we have

$$\begin{aligned}\mathcal{C}_V(t_1,t_2) &= \mathbb{E}\Big\{\big\{B_H[\theta(t_1+\Delta t)] - B_H[\theta(t_1)]\big\}\cdot\big\{B_H[\theta(t_2+\Delta t)] - B_H[\theta(t_2)]\big\}\Big\} = \\ &= \mathbb{E}\Big\{B_H[\theta(t_1+\Delta t)]\cdot B_H[\theta(t_2+\Delta t)] - B_H[\theta(t_1+\Delta t)]\cdot B_H[\theta(t_2)] - \\ &\quad - B_H[\theta(t_1)]\cdot B_H[\theta(t_2+\Delta t)] + B_H[\theta(t_1)]\cdot B_H[\theta(t_2)]\Big\}.\end{aligned}$$

Since by Definition 1.3.1 we have

$$\mathcal{C}_{B_H}(t,s) = \mathbb{E}\left[B_H(t)\cdot B_H(s)\right] = \frac{\sigma^2}{2}\cdot\Big(t^{2H} + s^{2H} - |t-s|^{2H}\Big),$$

the previous covariance function is equal to

$$\begin{aligned}\mathcal{C}_V(t_1,t_2) &= \frac{\sigma^2}{2}\cdot\Big[\theta(t_1+\Delta t)^{2H} + \theta(t_2+\Delta t)^{2H} - |\theta(t_1+\Delta t) - \theta(t_2+\Delta t)|^{2H}\Big] - \\ &\quad - \frac{\sigma^2}{2}\cdot\Big[\theta(t_1+\Delta t)^{2H} + \theta(t_2)^{2H} - |\theta(t_1+\Delta t) - \theta(t_2)|^{2H}\Big] - \\ &\quad - \frac{\sigma^2}{2}\cdot\Big[\theta(t_1)^{2H} + \theta(t_2+\Delta t)^{2H} - |\theta(t_1) - \theta(t_2+\Delta t)|^{2H}\Big] + \\ &\quad + \frac{\sigma^2}{2}\cdot\Big[\theta(t_1)^{2H} + \theta(t_2)^{2H} - |\theta(t_1) - \theta(t_2)|^{2H}\Big] = \\ &= \frac{\sigma^2}{2}\cdot\Big[|\theta(t_1+\Delta t) - \theta(t_2)|^{2H} - |\theta(t_1+\Delta t) - \theta(t_2+\Delta t)|^{2H} + \\ &\quad + |\theta(t_1) - \theta(t_2+\Delta t)|^{2H} - |\theta(t_1) - \theta(t_2)|^{2H}\Big].\end{aligned}$$

Since, due to law of total expectation, the covariance function can be written as

$$\begin{aligned}\mathcal{C}_V(t_1,t_2) = \mathbb{E}\Big\{\mathbb{E}\big\{&[X(t_1+\Delta t) - X(t_1)]\cdot[X(t_2+\Delta t) - X(t_2)]\;|\,\theta(t_1) = u_1, \\ &\theta(t_2) = u_2, \theta(t_1+\Delta t) = u_3, \theta(t_2+\Delta t) = u_4\big\}\Big\},\end{aligned}$$

the inner expectation is equal to

$$\frac{\sigma^2}{2}\cdot\left(|u_3-u_2|^{2H}-|u_3-u_4|^{2H}+|u_1-u_4|^{2H}-|u_1-u_2|^{2H}\right).$$

Reiterating the expectation, since by Definition 3.4.1

$$\mathbb{E}\left[|\theta(t)-\theta(s)|^q\right]=c(q)\cdot|t-s|^{\tau(q)+1},$$

we have

$$\begin{aligned}\mathcal{C}_V(t_1,t_2)=\frac{\sigma^2}{2}\cdot\Big[&\mathbb{E}\left(|u_3-u_2|^{2H}\right)-\mathbb{E}\left(|u_3-u_4|^{2H}\right)+\mathbb{E}\left(|u_1-u_4|^{2H}\right)-\\&-\mathbb{E}\left(|u_1-u_2|^{2H}\right)\Big]=\\=c_\theta(2H)\cdot\frac{\sigma^2}{2}\cdot\Big(&|t_1+\Delta t-t_2|^{\tau_\theta(2H)+1}-|t_1+\Delta t-t_2-\Delta t|^{\tau_\theta(2H)+1}+\\&+|t_1-t_2-\Delta t|^{\tau_\theta(2H)+1}-|t_1-t_2|^{\tau_\theta(2H)+1}\Big).\end{aligned}$$

With some arrangements, we find that the autocovariance of the incremental process is

$$c_\theta(2H)\cdot\frac{\sigma^2}{2}\cdot\left(|t_1-t_2+\Delta t|^{\tau_\theta(2H)+1}+|t_1-t_2-\Delta t|^{\tau_\theta(2H)+1}-2\cdot|t_1-t_2|^{\tau_\theta(2H)+1}\right)$$

□

If in addiction we postulate $t_1\geq t_2+\Delta t$, all the absolute values can be removed becoming

$$c_\theta(2H)\cdot\frac{\sigma^2}{2}\cdot\left[(t_1-t_2+\Delta t)^{\tau_\theta(2H)+1}+(t_1-t_2-\Delta t)^{\tau_\theta(2H)+1}-2\cdot(t_1-t_2)^{\tau_\theta(2H)+1}\right]$$

Consistently, if $H=\frac{1}{2}$ (that is a SBM), we find that the autocovariance function of the incremental process vanishes. In fact $\tau_\theta\left(2\cdot\frac{1}{2}\right)=\tau_\theta(1)=0$, and hence

$$\begin{aligned}\mathcal{C}_V(t_1,t_2)&=c_\theta(1)\cdot\frac{\sigma^2}{2}\cdot\left[(t_1-t_2+\Delta t)^1+(t_1-t_2-\Delta t)^1-2\cdot(t_1-t_2)^1\right]\\&=c_\theta(1)\cdot\frac{\sigma^2}{2}\cdot(t_1-t_2+\Delta t+t_1-t_2-\Delta t-2\cdot t_1+2\cdot t_2)=0.\end{aligned}$$

Thus the persistence of $\mathcal{C}_V$ is determined by the sign of $\tau_\theta(2H)$. Since the scaling function $\tau_\theta(q)$ has the same sign of $q-1$, when $H>\frac{1}{2}$ (the FBM has long-range dependence) we find that $\tau_\theta(2H)$ is positive, since $2H-1>0$, and then the MMAR has long memory[31]; on the other hand, if $H<\frac{1}{2}$, the

[31] Here the definition of long memory requires more attenction. As a matter of fact, unlike the fractional Brownian motion case, the MMAR is multifractal only on *bounded* intervals of times (and hence t cannot goes to infinity, if we want to preserve that property). In this case, long memory can be informally defined stating that the longest (apparent) cycle has approximatively the same length as the interval of definition. In this sense, the c.d.f. of a canonical measure has long memory.

converse is true.

We finally examine the dependence in the absolute values of return, which indicates that the MMAR exhibits persistence in volatility. It is convenient to define the *autocovariance function of the q-th power* of the increments, that is

$$\mathcal{A}_X(t_1,t_2,q) := \mathcal{C}_{|V|^q}(t_1,t_2) = \mathbb{E}\big[|V(t_1)|^q \cdot |V(t_2)|^q\big]\,,$$

where $V(t) := X(t+\Delta t) - X(t)$ is the incremental process. The following theorem is related to this function.

Theorem 3.5.5 *If $\{B_H(t)\}_{t\in[0,T]}$ is a fracional Brownian motion, and $\mathbb{E}\big[\theta(t)^{Hq}\big]$ is finite, the compound process satisfies*

$$\mathcal{A}_X(t_1,t_2,q) \geq \mathcal{A}_\theta(t_1,t_2,Hq)\cdot \mathbb{E}\big[|B_H(1)|^q\big]^2\,,$$

for all $0 \leq q < q_{crit}$, where the equality holds when $H = \frac{1}{2}$.

Proof: $\mathcal{A}_V(t_1,t_2,q)$ in terms of the return process $\{X(t)\}$ is given by

$$\begin{aligned}\mathcal{A}_V(t_1,t_2,q) &= \mathbb{E}\big[|X(t_1+\Delta t) - X(t_1)|^q \cdot |X(t_2+\Delta t) - X(t_2)|^q\big] = \\ &= \mathbb{E}\Big[\big|\,[X(t_1+\Delta t) - X(t_1)]\cdot[X(t_2+\Delta t) - X(t_2)]\,\big|^q\Big]\,.\end{aligned}$$

Using the law of total expectation, it can be rewritten as

$$\begin{aligned}\mathcal{A}_V(t_1,t_2,q) = \mathbb{E}\Big\{\mathbb{E}\Big\{\big|\,[X(t_1+\Delta t) - X(t_1)]\cdot[X(t_2+\Delta t) - X(t_2)]\,\big|^q \,\big|\, \theta(t_1)=u_1,\\ \theta(t_2)=u_2, \theta(t_1+\Delta t)=u_3, \theta(t_2+\Delta t)=u_4\Big\}\Big\}\,,\end{aligned}$$

and using the same procedure of the proof of Theorem 3.5.4, it can be written as

$$\mathcal{A}_V(t_1,t_2,q) = \mathbb{E}\Big\{\mathbb{E}\Big\{\big|\,[B_H(u_3) - B_H(u_1)]\cdot[B_H(u_4) - B_H(u_2)]\,\big|^q\Big\}\Big\}\,.$$

Assuming that the FBM has either independent $\big(H=\frac{1}{2}\big)$ or positively correlated $\big(H>\frac{1}{2}\big)$ increments, therefore this quantity can be bounden below by the following

$$\mathcal{A}_V(t_1,t_2,q) \geq \mathbb{E}\Big\{\mathbb{E}\Big\{\big|B_H(u_3) - B_H(u_1)\big|^q\Big\}\cdot\mathbb{E}\Big\{\big|B_H(u_4) - B_H(u_2)\big|^q\Big\}\Big\}\,.$$

Using (1.17), we find (using the independence between $\{B_H(t)\}$ and $\{\theta(t)\}$)

$$\begin{aligned}\mathcal{A}_V(t_1,t_2,q) &\geq \mathbb{E}\Big\{\Big[\mathbb{E}\big[|B_H(1)|^q\big]\cdot|u_3-u_1|^{Hq}\Big]\cdot\Big[\mathbb{E}\big[|B_H(1)|^q\big]\cdot|u_4-u_2|^{Hq}\Big]\Big\} = \\ &= \mathbb{E}\big[|B_H(1)|^q\big]^2\cdot\mathbb{E}\Big[|u_3-u_1|^{Hq}\cdot|u_4-u_2|^{Hq}\Big] = \\ &= \mathbb{E}\big[|B_H(1)|^q\big]^2\cdot\mathbb{E}\Big[|\theta(t_1+\Delta t)-\theta(t_1)|^{Hq}\cdot|\theta(t_2+\Delta t)-\theta(t_2)|^{Hq}\Big] = \\ &= \mathbb{E}\big[|B_H(1)|^q\big]^2\cdot\mathcal{A}_\theta(t_1,t_2,Hq)\,.\end{aligned}$$

□

A very surprising consequence of this theorem is that the return process has long memory in the absolute value of its increments, but if $H = \frac{1}{2}$ (that is a SBM), the returns display both *uncorrelated increments* and *persistence in volatility*.

The last theoretical property of the MMAR is about its multifractal spectrum.

Theorem 3.5.6 *Under hypotheses* **H1**, **H2**, **H3**, *the process* $X(t)$ *has multifractal spectrum* $f_X(\alpha)$ *equal to*

$$f_X(\alpha) = f_\theta\left(\frac{\alpha}{H}\right). \tag{3.18}$$

Proof: The absolute variation of the compound process $\{X(t)\}$ around the time t is given by

$$\left|X(t+\Delta t) - X(t)\right| = \left|B_H[\theta(t+\Delta t)] - B_H[\theta(t)]\right|.$$

As $\Delta t \to 0$, by virtue of Proposition 2.4.8 and Definition 3.4.2, we have

$$\left|B_H[\theta(t+\Delta t)] - B_H[\theta(t)]\right| \sim \left|\theta(t+\Delta t) - \theta(t)\right|^H \sim \left|\Delta t\right|^{H\cdot\alpha_\theta(t)}$$

where $\alpha_\theta(t)$ is the coarse Hölder exponent of θ at time t. The coarse Hölder exponent of X at t id therefore equal to $H \cdot \alpha_\theta(t)$.

Since by (3.14), $f_X(\alpha) = \dim_{\mathcal{H}}(T_\alpha)$, where T_α is the set where X has Hölder exponent α, by the previous asymptotic relation, that set is then identical to the set where θ has exponent $\frac{\alpha}{H}$. Hence, they have the same fractal dimension

$$f_X(\alpha) = f_\theta\left(\frac{\alpha}{H}\right).$$

□

This theorem shows that the MMAR is multifractal since the trading time is allowed to have a continuum of coarse Hölder exponents, thus transferring multifractality through compounding. Moreover, long memory has an interesting geometric interpretation: When the return process is multiscaling, several sets T_α have non-integer fractal dimension $f(\alpha) \in (0,1)$. Their elements necessarily cluster in certain regions of the interval of definition, so explaining the alternation of periods of large and small price changes. The set T_α is also statistically self-similar, in the sense that after proper rescaling, subsets of T_α have statistically the same relative placement of points than the original T_α. Therefore, the knowledge of T_α in one period owns important information of T_α in later periods. This property is another "face" of the long memory involved in the process.

Furthermore, since

$$f_\theta\left(\frac{\alpha}{H}\right) = \dim_{\mathcal{H}}\left(\left\{t \in [0,T] : \lim_{\Delta t \to 0} \frac{\log\theta\left([t, t+\Delta t]\right)}{\log\left(\Delta t\right)} = \frac{\alpha}{H}\right\}\right),$$

thus the volatility of X at time t depends on the local dimension of θ at t, and so the distribution of times of high and low activity in the share prices depends on the multifractal structure of θ. Since the latter is a random variable, the times of high and low activity are also random.

Statistical methods may be applied to real price data to estimate both α and the multifractal spectrum for θ. In fact, due to (3.14) with some arrangements, we have

$$f_X\left(\alpha\right) = \dim_{\mathcal{H}}\left(\left\{t \in [0,T] : \limsup_{\Delta t \to 0} \frac{\log|X(t+\Delta t) - X(t)|}{\log|\Delta t|} = \alpha\right\}\right).$$

3.6 Noah and Joseph effects revisited

In this final section we again take into account the former models used for describing price changes (that is BM, FBM, and LSM), in order to compare their properties with those of MMAR. Moreover, we will see that the model we introduced in this chapter is the only one (among those discussed in this work) able to contemplate both Noah effect and Joseph effect contemporarily, with other suitable properties.

First of all, Theorem 3.5.3 states that when $H = \frac{1}{2}$ the process describing asset returns is a martingale, being so the future unpredictable from the knowledge of past prices. At the same time, Theorem 3.5.5 guarantees long memory in the absolute value of returns. That is, there is not only *time heterogeneity* in the size of log-returns, but also that heterogeneity is present on *any time scale* at which we decide to investigate the data, that is they may be daily, weekly, monthly, or yearly returns[32].

Moreover, the MMAR incorporates important features observed in financial time series, including long tails and invariance under scaling. Multifractality is defined by a set of restrictions on the process' moments as the time scale of observation change (that is the parameters of the model vary, but the structure of it does not modify). Furthermore, it is integrated in the model through *trading time*, a random distorsion of clock time that leads to strong changes in volatility.

In addiction, the MMAR allows the possibility for returns to be uncorrelated, but does not impose it: The model has got enough flexibility to

[32]The simultaneous presence of the martingale property associated with a memory of any sort might be surprising to many. However, the MMAR is not the only one model to have provided for it: Several authors have developed the so called FIGARCH models, which present both unpredictability of the future and memory. Nevertheless, the MMAR is the only one model to own that property, being scale-invariant as well.

satisfy the martingale property in some cases and and long memory in its increments. But, perhaps, the main advantage of this model over alternatives is the property of scale-invariance. Because of this property, the whole information contained at different frequencies can be used to identify and test the model.

The following table summarizes what just stated, with the convenience to report part of Table 1.1 allowing the comparison with the SBM, FBM, and LSM.

Table 3.1: Comparison of the MMAR with the former models.

	SBM	FBM	LSM	MMAR
Distribution of increments	Gaussian	Gaussian	L-stable	Gaussian
Independence of increments	✓	×	✓	✓[d]
Stationarity of increments	✓	✓	✓	✓
Mean of the process	0	0	∞ or 0^a	0
Covariance of the process	finite	finite	infinite	finite
Semi-martingale	✓	×	✓	✓[d]
Self-affinity/uniscaling	✓	✓	✓[b]	✓
Multiscaling	×	×	×	✓
Long-range dependence	×	✓[c]	×	✓[c]
Heaviness of the tails	×	×	✓	✓[e]

[a] The mean of the process is infinite if $0 < \alpha \leq 1$ and zero if $1 < \alpha < 2$.

[b] The process is self-affine if and only if $1 < \alpha < 2$.

[c] The process has long-range dependence if and only if $\frac{1}{2} < H < 1$.

[d] The process has independent increments and is a martingale if and only if $H = \frac{1}{2}$.

[e] Tails become heavy for those q-th absolute moments with $q \geq q_{crit}$.

The compatibility with the martingale property of returns and the long memory are given by Theorem 3.5.3 and Theorem 3.5.5 respectively. Oth-

erwise, scale-invariance and multiscaling are consequence of the definition of multifractal processes. They respectively correspond to the time-invariance and non-linearity of the scaling function $\tau(q)$. Moreover, the multiscaling properties of the MMAR contrast with the unique scale contained in all the previous financial models.

The main disadvantage of the MMAR is the dearth of applicable statistical methods, although the authors of it – Mandelbrot, Calvet and Fisher – introduced few innovative tools able to be used for models being both time-invariant and scale-invariant.

Furthermore, the connection with the multifractal formalism and the multifractal spectrum is also crucial. As a matter of fact, the former, $f(\alpha)$, can be seen as a renormalized density obtained as the limits of histograms. In an alternative, $f(\alpha)$ is viewed as the fractal dimension of the set of instants T_α with local Hölder exponent α. The statistical self-similarity of the set T_α is closely related to the long memory. In addiction, for a large class of multifractal measures, the spectrum can be explicitly derived from Large Deviation Theory through the Legendre transform. This allow the researcher to relate an empirical estimate of the spectrum back to a particular construction of the multifractal. On this direction goes the other paper of the authors [9].

The consequences on the MMAR can be so interpreted. The heterogeneity of the local scales along the price process is entirely given by the trading time $\theta(t)$. Moreover, the multifractal spectrum of the *price* is derived from the of the *trading time* by a very simple transformation.

Other practical aspects are of great interest to financial economics. Those are mainly the consequences of using a multifractal process and multifractality as underlying assumptions: The clear non-classic Brownian (mis)behaviour of the multifractal processes has enormous implication over the risk-adverse investors' decisions and the pricing of derivatives.

Chapter 4

Case study: The best model for ENI.MI, FNC.MI, G.MI, and UCG.MI

In the last chapter we are interested in estimating the parameters of the stochastic processes describing four blue-chips traded on the Italian stock market. In particular we have selected four shares whose companies belong to different industries.

Stock code	Corporation	Industry	Capitalization	ADV
ENI.MI	Eni	Energy	60.03	11.14
FNC.MI	Finmeccanica	Aeronautics	2.21	4.05
G.MI	Assicurazioni Generali	Insurance	22.45	5.84
UCG.MI	Unicredit	Bank	23.37	76.32

Capitalization is referred as billions of euros (€) and Average Daily Volumes (ADV) are expressed as millions of traded units (in addiction, the average has been computed on the period May 2013 - July 2013).

The first aim of this chapter is to show how different can be the outcomes – for option pricing, risk management, portfolio optimization, etc – according to the different hypotheses on the underlying stochastic process. Since market agent are interest in two main cornerstones, *return* and *risk* (having in mind that it may be measured in a lot of different ways), we will describe the assets return's process though the four models previously introduced, and show how risk varies.

The second aim is to give a honest and concrete evaluation of models ability and flaws in describing the stocks' movements. In order to do that, we will firstly estimate the required parameters of the different models (namely

BM, FBM, LSM, and MMAR). About the estimation techniques, we will give only a sketch. For a further in-depth analysis, we refer to the quotes in the text. Secondly, according to those estimates, we will simulate the stock's price paths. Also a "visual" comparison will be useful to see the discrepancies among theory and reality.

Eventually, the last section of the chapter will be devoted to a concrete application in *option pricing*. There, the different ways of modelling and measuring risk will reveal the danger involved in using a *mild* model rather than a *wild* one.

The financial time series of the stock prices are taken from 01/01/2008 to 31/07/2013, and will be analyzed with the free software  . As a matter of fact, several libraries and finance-specific function already exist (to download the data we will use the function `get.hist.quote`, which allow to use the Yahoo!Finance database). Moreover, there are several packages useful in order to quantify long-range dependence parameters (like the Hurst exponent H), and two dedicated libraries to estimate and simulate stable random variables.

Log-returns are again defined in the following way

$$X(t, \Delta t) := X(t) - X(t - \Delta t) = \ln [S(t)] - \ln [S(t - \Delta t)] \ ,$$

where $X(t) := \ln[S(t)] - \ln[S(t_0)]$ and $\{S(t)\}$ is the stochastic process describing the stock price at time t, with $t_0 \leq t - \Delta t < t \leq T$. In our descriptive analysis we set $t_0 = 01/01/2008$ and $T = 31/07/2013$. In the last section of the chapter, which will be devoted to option pricing outcomes based on the following models.

Moreover, in order to implement the MMAR, we will need to work with different frequencies (see Section 4.2). Of course the all the records are based only on working days.

4.1 Estimating the parameters of BM, FBM and LSM

The first model of asset return we are going to deal with is Brownian motion (BM). Due to Option Pricing Theory, it is used to state this property in terms of price diffusion equation, that is represented by the following *stochastic differential equation* (hereinafter named by SDE)

$$\frac{\mathrm{d}S(t)}{S(t)} = \mu \cdot \mathrm{d}t + \sigma \cdot \mathrm{d}B(t) \tag{4.1}$$

where $\{S(t)\}$ is the stochastic process describing the stock price at time t, $\mathrm{d}B(t)$ is the formal differential of a SBM (see section 1.2), μ is the *drift*

parameter, and σ is the so-called *diffusion coefficient*. Since it is well-know that the previous SDE has as solution (see [6])

$$S(t) = S_0 \cdot e^{\left(\mu - \frac{\sigma^2}{2}\right)\cdot t + \sigma \cdot B(t)}, \tag{4.2}$$

having set $S(t_0) = S_0$ as initial condition[1]. This equation will be used later in order to simulate the possible paths of the stocks' prices, and then implement the Monte Carlo simulations. Rearranging the last expression, we find

$$\ln\left[S(t)\right] = \ln\left(S_0\right) + \left(\mu - \frac{\sigma^2}{2}\right) \cdot t + \sigma \cdot B(t)\,,$$

that is $\ln\left[S(t)\right] - \ln\left(S_0\right)$ is a Brownian motion in the sense of (1.9), with drift and diffusion coefficient equal to $\mu - \frac{\sigma^2}{2}$ and σ respectively.

Taking two different time instants t and $t - \Delta t$, and subtracting the corresponding two solutions we find the log-return between $t - \Delta t$ and t, that is

$$X(t, \Delta t) = \ln\left[S(t)\right] - \ln\left[S(t - \Delta t)\right] = \left(\mu - \frac{\sigma^2}{2}\right) \cdot \Delta t + \sigma \cdot \Big[B(t) - B(t - \Delta t)\Big]\,.$$

If we set $\Delta t = 1$ (day), we get

$$X(t, 1) = \left(\mu - \frac{\sigma^2}{2}\right) + \sigma \cdot \Big[B(t) - B(t-1)\Big]$$

which is the discrete version of a Gaussian white noise (with a non-zero intercept and a diffusion coefficient).

It is very useful to compute the expectation and variance of the last quantity, since through the following results we will be able to estimate the necessary parameter to simulate equation (4.1). Using property **(c)** involved in Definition 1.1.1, they are respectively

$$\begin{aligned}
\mathbb{E}\big[X(t,1)\big] &= \mathbb{E}\left\{\left(\mu - \frac{\sigma^2}{2}\right) + \sigma \cdot \Big[B(t) - B(t-1)\Big]\right\} = \\
&= \mu - \frac{\sigma^2}{2} + \sigma \cdot \underbrace{\mathbb{E}\Big[B(t) - B(t-1)\Big]}_{=0} = \\
&= \mu - \frac{\sigma^2}{2}\,,
\end{aligned}$$

[1] Such a process is commonly addressed as a *Geometric Brownian motion* (GBM).

and

$$\begin{aligned}\mathbb{V}\big[X(t,1)\big] &= \mathbb{V}\left\{\left(\mu - \frac{\sigma^2}{2}\right) + \sigma \cdot \Big[B(t) - B(t-1)\Big]\right\} = \\ &= \sigma^2 \cdot \underbrace{\mathbb{V}\Big[B(t) - B(t-1)\Big]}_{=t-(t-1)} = \\ &= \sigma^2 \,.\end{aligned}$$

Hence, given the time series of the daily log-returns

$$x_t := \ln s_t - \ln s_{t-1}$$

with $t_0 \leq t-1 < t \leq T$, the diffusion coefficient can be estimated through the unbiased estimator of the variance of the daily returns, that is[2]

$$\hat{\sigma} = \sqrt{\widehat{\mathbb{V}\big[X(t,1)\big]}} = \sqrt{\frac{\sum_{t=1}^{T}(x_t - \bar{x})^2}{T-1}} \tag{4.3}$$

where $\bar{x} = \widehat{\mathbb{E}\big[X(t,1)\big]} = \frac{\sum_{t=1}^{T} x_t}{T}$ is the sample mean. Moreover, a very easy way to estimate the drift parameter of equation (4.1) is achieved using the previous results, that is

$$\hat{\mu} = \bar{x} + \frac{\hat{\sigma}^2}{2} \,. \tag{4.4}$$

Through these estimates we compute the drift parameter and the diffusion coefficient for all four stock. The  codes, both for BM and for all the other models discussed in the following dissertation, can be found in Appendix D. The results, based on daily records, are reported in Table 4.1.

Table 4.1: Estimates for the BM models.

Stock code	$\hat{\mu}$	$\hat{\sigma}$
ENI.MI	-0.000076	0.020308
FNC.MI	-0.000836	0.026484
G.MI	-0.000292	0.020407
UCG.MI	-0.000767	0.042892

We can see that the three first stocks have exhibited almost the same daily volatility (intended in units of standard deviation). Only UCG.MI has got a higher/double risk (that is likely due to the recent downfalls after bank industry crisis). On the other hand, all have shown a very close-to-zero

[2] To simplify the notation, we have set $t_0 = 0, t_1 = 1, \ldots$

negative drift, confirming the overall bad performances of the Italian (and European) stock market during the last years.

Via these estimates and using (4.1) we can replicate the past price evolution of the four stocks. The figure below illustrate it, using daily log-return rather than price: If we were plotting real price against simulated one, we could be mislead. Otherwise, returns show properly the inappropriateness of BM as a model of asset return.

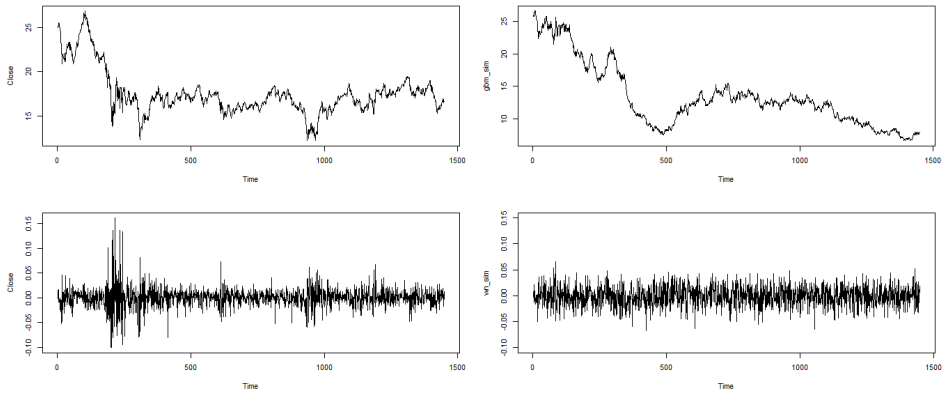


Figure 4.1: Real ENI's data on the left side; simulation to the right.

Of course, the same is valid for the other three stocks (we avoid to present also the prices' plots in order to focus only on log-returns).

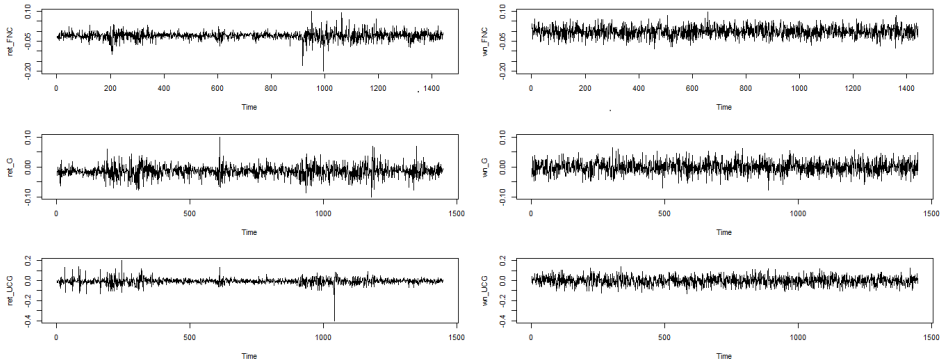


Figure 4.2: Real data on the left side; simulations to the right.

We might dispassionately conclude that Brownian motion is not an accurate stochastic process in order to describe log-prices. The consequences of employing such a model will be analysed and disclosed in the last section of this chapter, when we will price the American call options written on those stocks. As a matter of fact, the very *mild* variability of BM tends to underestimate the *risk* (here embodied by the diffusion coefficient σ), a crucial

feature involved into option pricing.

As shown by the the densities of Figure 4.3, BM does not (or rather, *can not*) take into account the information contained inside the tail of the actual distributions.

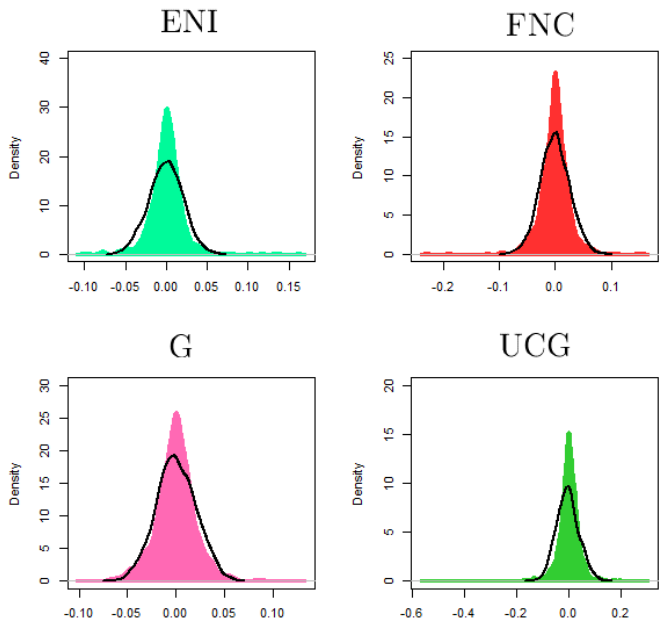


Figure 4.3: Densities of real data vs. densities generated by BM.

Now we will take into account the second stochastic process discussed in this work, namely the fractional Brownian motion (FBM). The simplest way to apply it to financial prices is to change (4.1) into

$$S(t) = S_0 \cdot e^{\left(\mu - \frac{\sigma^2}{2}\right)\cdot t + \sigma \cdot B_H(t)}, \tag{4.5}$$

where we have just replaced a Brownian motion $B(t)$ with a fractional Brownian motion $B_H(t)$. Hence, using a FBM we are requested two estimates three parameters instead of only two. The third one is obviously the Hurst exponent H.

The procedure to estimate μ and σ fortunately remain the same. In fact, using daily log-return and by virtue of (1.17) we find

$$\begin{aligned}
\mathbb{E}\big[X(t,1)\big] &= \mathbb{E}\left\{\left(\mu - \frac{\sigma^2}{2}\right) + \sigma \cdot \Big[B_H(t) - B_H(t-1)\Big]\right\} = \\
&= \mu - \frac{\sigma^2}{2} + \sigma \cdot \underbrace{\mathbb{E}\Big[B_H(t) - B_H(t-1)\Big]}_{=0} = \\
&= \mu - \frac{\sigma^2}{2},
\end{aligned}$$

and

$$\begin{aligned}\mathbb{V}\big[X(t,1)\big] &= \mathbb{V}\left\{\left(\mu - \frac{\sigma^2}{2}\right) + \sigma \cdot \Big[B_H(t) - B_H(t-1)\Big]\right\} = \\ &= \sigma^2 \cdot \underbrace{\mathbb{V}\Big[B(t) - B(t-1)\Big]}_{=[t-(t-1)]^{2H}} = \\ &= \sigma^2 \,.\end{aligned}$$

Thus the quantity $\hat{\mu}$ and $\hat{\sigma}$ remain unchanged. The only additional complexity is to estimate the self-affinity parameter H. In Section 1.7 we have already introduced a way of estimating it due to Hurst itself and improved for the first time by Mandelbrot, that is *Rescaled Range Analysis*, usually shortened as *R/S Analysis*. Particularly, the goal will be reached via the function `rsFit`. Even though already introduced, hereafter we will display its essential features.

Let $x_1, x_2, ..., x_T$ be the values of T successive daily log-return. We start computing these return for the different intervals k until the entire time series of T days. Then, for every interval, it is computed the difference between the corresponding rate and the sample mean. Then the cumulative values are taken, that is

$$\sum_{t=1}^{k}\left(x_t - \frac{1}{T}\sum_{t=1}^{T} x_t\right) = \sum_{t=1}^{k} x_t - \frac{k}{T}\sum_{t=1}^{T} x_t$$

with $k = 1, 2, ..., T$. Considered the whole sample, we will have a maximum and a minimum difference. Subtracting these two values we have an estimate of the *sample range* from the highest to the least peak of the deviation of the cumulative return, that is

$$\mathcal{R}_k := \max_{1\le k\le T}\left(\sum_{t=1}^{k} x_t - \frac{k}{T}\sum_{t=1}^{T} x_t\right) - \min_{1\le k\le T}\left(\sum_{t=1}^{k} x_t - \frac{k}{T}\sum_{t=1}^{T} x_t\right) .$$

Furthermore the empirical *standard deviation* is

$$\mathcal{S}_k := \sqrt{\frac{1}{k}\sum_{t=1}^{k}(x_t - \bar{x}_k)^2}\,,$$

where $\bar{x}_k = \frac{\sum_{t=1}^{k} x_t}{k}$. Hurst and Mandelbrot showed that as $k \to +\infty$

$$\frac{\mathcal{R}_k}{\mathcal{S}_k} \sim C \cdot k^H$$

for a positive constant C and $0 < H < 1$, where the latter is the usual Hurst exponent. Setting $a_k = \lim_{k\to\infty} \frac{\mathcal{R}_k}{\mathcal{S}_k}$, the estimate of H can be obtained as the slope of the log-log plot as k varies. As a matter of fact

$$\ln(a_k) = \ln(C) + H \cdot \ln(k)\,.$$

Using the ordinary least square estimate, we find

$$\hat{H} = \frac{\sum_{k=1}^{T}\left(\ln(k) - \sum_{k=1}^{T}\frac{\ln(k)}{T}\right)\cdot\left(\ln\left(a_k\right) - \sum_{k=1}^{T}\frac{\ln(a_k)}{T}\right)}{\sum_{k=1}^{T}\left(\ln(k) - \sum_{k=1}^{T}\frac{\ln(k)}{T}\right)^2}. \tag{4.6}$$

If the time series is generated by a Brownian motion, it has $H = \frac{1}{2}$. If the sample range is wider than the normal case ($H > \frac{1}{2}$), data are persistent; conversely, if it is more limited ($H < \frac{1}{2}$), data are anti-persistent and always tend to change direction.

The results of that procedure are the one of Table 4.2.

Table 4.2: Estimates for the FBM models.

Stock code	$\hat{H}$
ENI.MI	0.560669
FNC.MI	0.621358
G.MI	0.595978
UCG.MI	0.577973

Hence, all the stocks possesses a form of long memory and persistency.

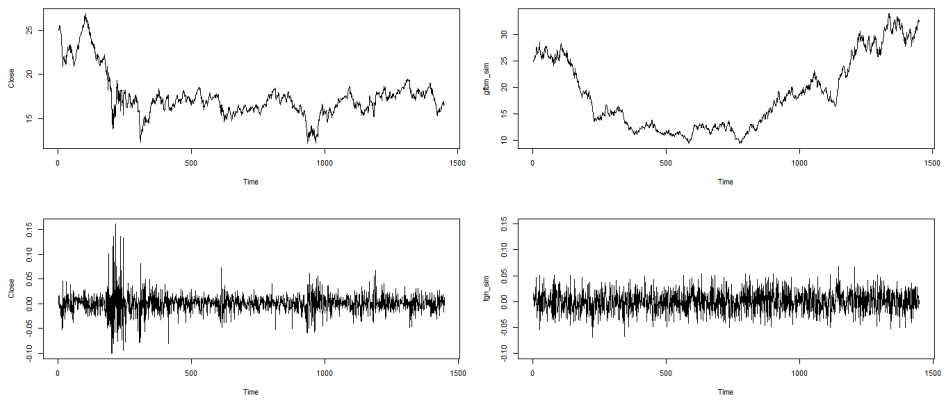


Figure 4.4: Real ENI's data on the left side; simulation to the right.

Despite the fact that FBM takes into account that form of memory – totally absent using BM as the driver of the log-prices – it is still too mild, being so unappropriate in order to describe the phenomenon of price changes.

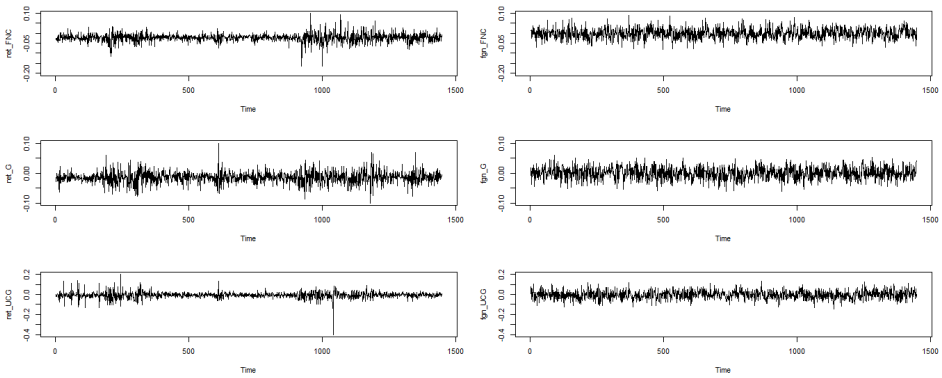


Figure 4.5: Real data on the left side; simulations to the right.

The estimated densities in Figure 4.6 confirm that inappropriateness. Moreover, the fact of having introduced and estimated the memory-parameter H seems not to have improved nor described better real data.

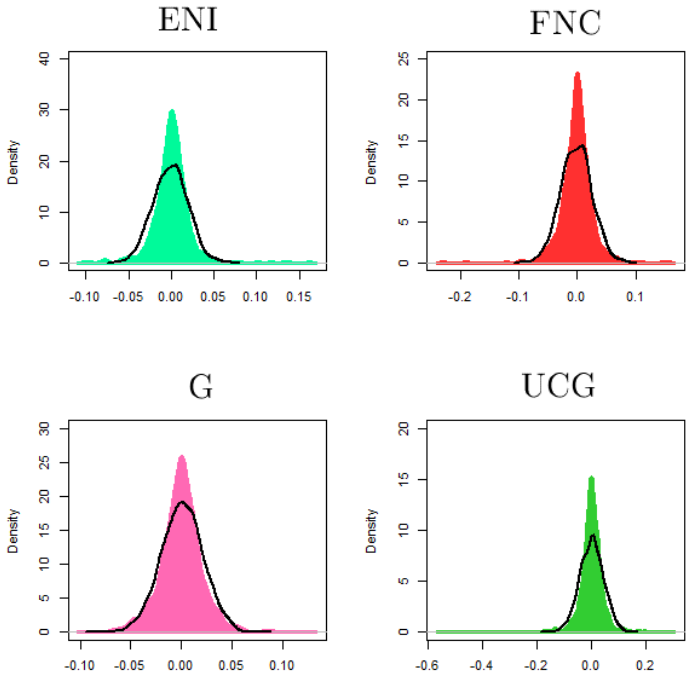


Figure 4.6: Densities of real data vs. densities generated by FBM.

The last part of this section is devoted to simulating a LSM as the model describing asset log-returns. Since they are supposed to be distributed as a stable random variable (which is fully described by *four* parameters), the

estimate procedure can become far more complicated, basically depending on the method chosen. In the following, we will opt for one of the simplest one, which the function `stableFit` is based on.

That technique is named *Quantile estimation* and was firstly introduced by Fama and Roll in [16] for a symmetric zero-centred stable law (that is $\beta = 0$ and $\delta = 0$), with the shape parameter $\alpha > 1$. Later, McCulloch in [30] generalized and improved their method. He analyzed stable law quantiles and provided consistent estimators of all four stable parameters, with the restriction $\alpha \geq 0.6$, while retaining the computational simplicity of Fama and Roll's method. Since we suppose, *a posteriori*, that $1 < \alpha \leq 2$, the restriction will not affect our purpose[3]. We will give just a hint of this estimation procedure.

Given the sample $x_1, x_2, ..., x_T$ of T successive daily log-return, let we define

$$\zeta_\alpha = \frac{x_{0.95} - x_{0.05}}{x_{0.75} - x_{0.25}}$$

which is both independent of both γ and δ, and where x_f denotes the f-th population quantile, so that $\mathcal{S}(\alpha, \beta, \gamma, \delta)(x_f) = f$. Let $\widehat{\zeta_\alpha}$ be the corresponding sample value, being a consistent estimator of ζ_α as well. Now define

$$\zeta_\beta = \frac{x_{0.95} + x_{0.05} - 2 \cdot x_{0.50}}{x_{0.75} - x_{0.25}}$$

and let $\widehat{\zeta_\beta}$ be the corresponding sample value, being again a consistent estimator of ζ_β.

It can be proved that statistics ζ_α and ζ_β are functions of index of stability α and the skewness parameter β only. Moreover, this relationship may be inverted and the parameters α and β may be viewed as functions of ζ_α and ζ_β:

$$\alpha = \psi_1(\zeta_\alpha, \zeta_\beta) \qquad \text{and} \qquad \beta = \psi_2(\zeta_\alpha, \zeta_\beta) \ . \tag{4.7}$$

Substituting ζ_α and ζ_β by their sample values and applying linear interpolation between values found in tables provided by McCulloch in [30] yields estimators $\hat{\alpha}$ and $\hat{\beta}$.

Scale and location parameters, γ and δ, can be estimated in a similar way. However, due to the discontinuity of the characteristic function for $\alpha = 1$ (with $\beta \neq 0$), this procedure is much more complicated. For an in-depth analysis of that topic, see the original work [30].

Using our data, we have found the values of Table 4.3.

[3] The restriction on the shape parameter $\alpha \in (1, 2]$ is imposed by the necessity of dealing with a self-affine stochastic process, being that feature noticed in financial time series. See Theorem 1.6.12.

Table 4.3: Estimates for the LSM models.

Stock code	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\gamma}$	$\hat{\delta}$
ENI.MI	1.521	-0.168	0.009590	-0.000282
FNC.MI	1.489	-0.143	0.013138	-0.001186
G.MI	1.519	-0.054	0.011258	-0.000501
UCG.MI	1.464	-0.110	0.019243	-0.001687

As expected all the estimations of the index of stability are inside the interval, allowing so self-affinity[4].

In order to generate a process based on LSM for log-prices, we need to remind that a stable r.v. $X \overset{d}{\sim} \mathcal{S}(\alpha, \beta, \gamma, \delta)$ collapses towards a gaussian one $X \overset{d}{\sim} \mathcal{N}(m, \sigma^2)$ when $\alpha = 2$. Moreover, it follows that (see Section 1.5)

$$\gamma^2 = \frac{\sigma^2}{2} \quad \text{and} \quad \delta = m\,.$$

Since gaussianity is just a special case of stability (and hence BM is just a special case of LSM), we need to take into account such a process

$$\ln\left[S(t)\right] = \ln\left(S_0\right) + \left(\mu - \gamma^2\right) \cdot t + \sqrt{2} \cdot \gamma \cdot M(t)\,,$$

where $M(t)$ is a standard Lévy-stable motion (see Definition 1.6.11) and μ is similarly defined as in (4.4). This is due because $\alpha \in (1, 2]$, thus allowing the finiteness of the first moment. That is[5]

$$\begin{aligned}
\mathbb{E}\big[X(t,1)\big] &= \mathbb{E}\left\{\left(\mu - \gamma^2\right) + \sqrt{2} \cdot \gamma \cdot \Big[M(t) - M(t-1)\Big]\right\} = \\
&= \mu - \gamma^2 + \sqrt{2} \cdot \gamma \cdot \underbrace{\mathbb{E}\Big[M(t) - M(t-1)\Big]}_{=0} = \\
&= \mu - \gamma^2\,.
\end{aligned}$$

[4]Given a random variable $X \overset{d}{\sim} \mathcal{S}(\alpha, \beta, \gamma, \delta)$ with $\alpha \in (1, 2]$, by virtue of (1.36), we know that $\delta = \mathbb{E}(X)$. Then we have

$$\hat{\delta} = \widehat{\mathbb{E}(X)} = \frac{\sum_{t=1}^{T} x_t}{T}\,.$$

However, if we look at `stableFit`'s output, we will se that estimate of the location parameter is different from the sample mean. This is due to the fact that the function `stableFit` uses the Nolah's parameterization instead of the one given in (1.31). It changes only in the way of defining δ, and is usually employed because of its computational ease. Anyway, the values reported in the table are the ones provided though sample mean estimation. For a better understanding, see [33].

[5]Since it is now supposed $X(t,1) \overset{d}{\sim} \mathcal{S}(\alpha, \beta, \gamma, \delta)$ with $\alpha > 1$, we have

$$\hat{\mu} = \hat{\delta} + \hat{\gamma}^2\,.$$

Thus the stochastic process describing price evolution can be written as

$$S(t) = S_0 \cdot e^{\left(\mu-\gamma^2\right)\cdot t+\sqrt{2}\cdot\gamma\cdot M(t)}\,. \tag{4.8}$$

According to the estimates given in the previous table, we have generated the following simulations.

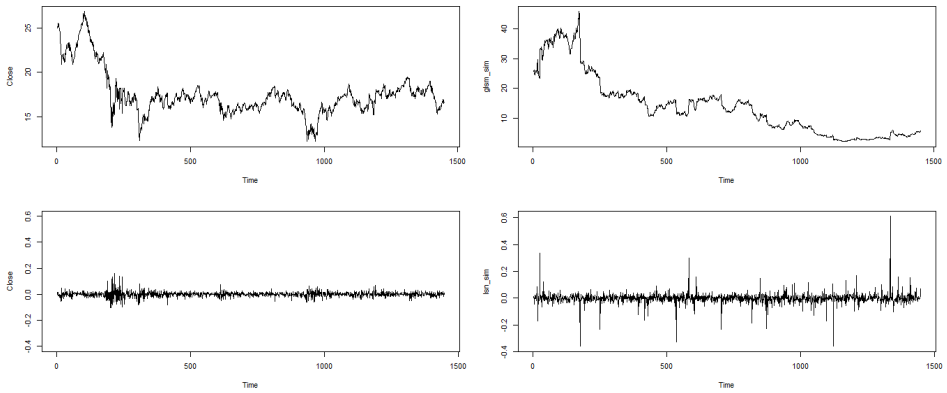


Figure 4.7: Real ENI's data on the left side; simulation to the right.

The figures show that using a LSM as stochastic process to describe log-prices is visually more accurate than the previous two cases. However, is far from being effective in representing actual financial data. It owns no memory and *heavy* variation are too large.

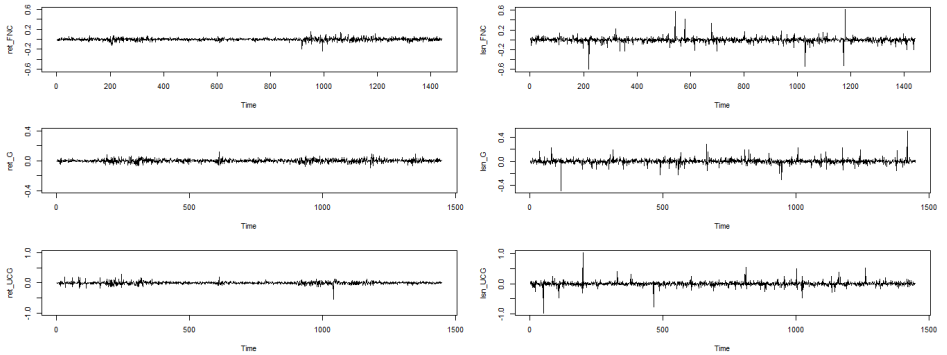


Figure 4.8: Real data on the left side; simulations to the right.

In any case, we want to highlight once more that both FBM and LSM as processes for log-prices and the corresponding "geometric versions" (4.5) and (4.8) for prices are both able to catch further information about real data. However, nor long-range dependence or heavy tails really can separately explain actual price fluctuations much better than Brownian motion. Indeed, the complex procedures of estimation involved in FBM and LSM

are not justified by inadequacy of the fitting, and might produce even vary biased outcomes.

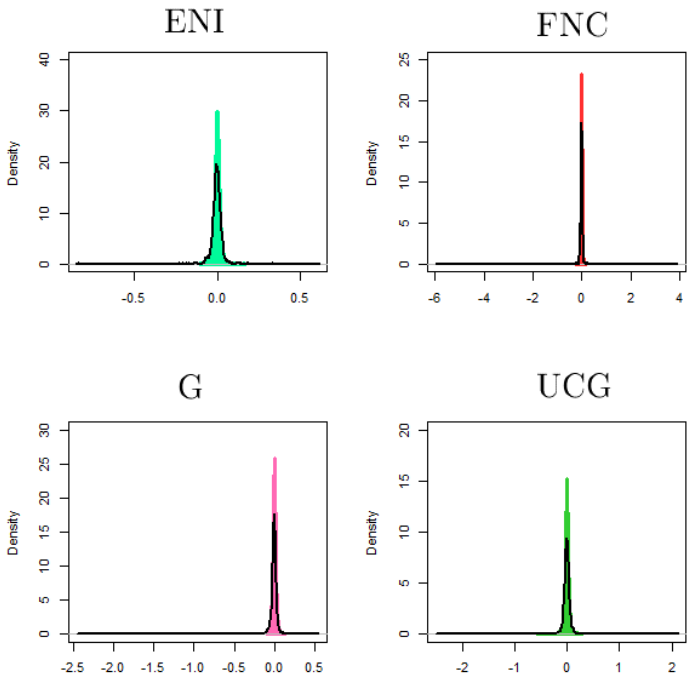


Figure 4.9: Densities of real data vs. densities generated by LSM.

The following section aim at using the MMAR as stochastic process of log-return. There are only few attempt in literature which used a multifractal process to describe financial data. Subsequently, there are no pre-set function able to estimate the quantities required to simulate the MMAR. Furthermore, there exist only few estimation techniques to infer multifractality upon financial data, whose rigorousness is still controversial.

4.2 Estimating the parameters of MMAR

In order to be able to simulate and replicate properly data whose underlying model is the MMAR we will have to estimate four quantities at least. These are the Hurst exponent H of the compounding process, the scaling function $\tau(q)$, the multiplicative prefactor $c(q)$, and, if the scaling function become null in more than one point, the critical exponent q_{crit} (for the explanation of the reason being, see Appendix C.1). As a matter of fact, defining

the log-returns as in the previous section

$$X(t,\Delta t) := X(t) - X(t-\Delta t) = \ln\left[S(t)\right] - \ln\left[S(t-\Delta t)\right] ,$$

we are now assuming

$$X(t) = B_H[\theta(t)]$$

where $\theta(t)$ is a multifractal process deforming physical time. By the definition of multifractal process, it implies

$$\mathbb{E}\big[|\theta(t)-\theta(s)|^q\big] = c_\theta(q)\cdot|t-s|^{\tau_\theta(q)+1} ,$$

where the expectation of the absolute q-th moment is assumed to be finite only for those $q \in [0, q_{crit})$ (see Appendix C.1).

Since, by virtue of Theorem 3.5.2, multifractality is extended to the process of asset return, that is

$$\mathbb{E}\big[|X(t)-X(s)|^q\big] = c_X(q)\cdot|t-s|^{\tau_X(q)+1} ,$$

with $c_X(q) = \mathbb{E}\big[|B_H(1)|^q\big]\cdot c_\theta(Hq)$ and $\tau_X(q) = \tau_\theta(Hq)$, all the required quantities can be inferred by the data. However, we will see that the estimation techniques are almost non-standard (and heuristic in some cases). All the methods describe hereinafter have been developed using those introduced by Mandelbrot, Calvet and Fisher in [9].

Let $\{S(t)\}_{t\in[0,T]}$ be the stochastic process describing the stock price at time t. Dividing $[0,T]$ into N intervals of length Δt, define the *partition function*

$$P(q,\Delta t) := \sum_{i=1}^{N}\big|X(i\Delta t,\Delta t)\big|^q . \tag{4.9}$$

This entity reveal itself to be crucial for those estimates we are dealing with. As a matter of fact, computing its expectation, and due to Definition 3.4.1, we find

$$\begin{aligned}
\mathbb{E}\big[P(q,\Delta t)\big] &= \mathbb{E}\left[\sum_{i=1}^{N}\big|X(i\Delta t)-X(i\Delta t-\Delta t)\big|^q\right] = \\
&= \sum_{i=1}^{N}\mathbb{E}\big[|X(i\Delta t)-X(i\Delta t-\Delta t)|^q\big] = \sum_{i=1}^{N} c_X(q)\cdot\Delta t^{\tau_X(q)+1} = \\
&= N\cdot c_X(q)\cdot\Delta t^{\tau_X(q)+1} = (N\cdot\Delta t)\cdot c_X(q)\cdot\Delta t^{\tau_X(q)} = \\
&= T\cdot c_X(q)\cdot\Delta t^{\tau_X(q)} .
\end{aligned}$$

Eventually, taking the logarithms of both sides, we get a linear regression model whose parameters (slope and intercept) may be estimated via OLS.

That is[6]

$$\ln\Big[\mathbb{E}\big[P(q,\Delta t)\big]\Big] - \ln(T) = \underbrace{\ln\big[c_X(q)\big]}_{\text{intercept}} + \underbrace{\tau_X(q)}_{\text{slope}} \cdot \ln(\Delta t)\,. \tag{4.10}$$

when the q-th moment exists finite[7].

The application of (4.10) is immediate: We plot $\ln[P(q,\Delta t)]$ against $\ln(\Delta t)$ for various values of q and various values of Δt. The linearity of these plots for given q is proposed in [9] as a *test* of the MMAR. The slope and intercept of the lines, estimated by OLS regression, give an estimate of the scaling function $\widehat{\tau_X(q)}$ and of the multiplicative prefactor $\widehat{c_X(q)}$ respectively. Furthermore, the estimated scaling function can then be transformed into an estimated multifractal spectrum $\widehat{f_X(\alpha)}$ using Legendre transform (see Appendix C.1).

The fact of employing $\ln[P(q,\Delta t)]$, in order to make the estimation procedures functional, means to use

$$\mathbb{E}\Big[\ln\big[P(q,\Delta t)\big]\Big] - \ln(T) = \ln\big[c_X(q)\big] + \tau_X(q)\cdot\ln(\Delta t)\,. \tag{4.11}$$

instead of (4.10). While inversion of logarithms and expectations will produce small sample biases in estimation[8], these do not prevent transmission of linearity in (4.10) to approximate linearity in the sample partition function.

Despite the ease of equation (4.11), consistency[9] of the estimated slopes and intercepts have only been conjectured. This topis, along with inference methods, remains an open and potentially fruitful area for further theoretical developments.

---

[6] We highlight that the regression equation provided by [9] is similar to (4.10) but incorrect.

[7] Since $\mathbb{E}\big[|\theta(t)-\theta(s)|^q\big] < \infty$ if and only if $q \in [0, q_{crit})$, and $c_X(q) = \mathbb{E}\big[|B_H(1)|^q\big] \cdot c_\theta(Hq)$ (the function $c(q)$ is the one responsible for the finiteness/infiniteness), we find

$$\mathbb{E}\big[|X(t)-X(s)|^q\big] < \infty$$

if and only if $q \in \big[0, \frac{q_{crit}}{H}\big)$.

[8] By Jensen's inequality, we have

$$\ln\Big[\mathbb{E}\big[P(q,\Delta t)\big]\Big] \geq \mathbb{E}\Big[\ln\big[P(q,\Delta t)\big]\Big]\,.$$

Hence equation (4.11) yields errors with negative expectation. Even in [9], the authors underline that the effects over estimations depends on both magnitude of the bias and how the bias varies with Δt. However, both problems should be correctable when taken into account for simulation purposes.

[9] Let $X_1, X_2, ..., X_n$ a sequence of random variables, such that $X \overset{d}{\sim} F_X(x;\theta)$, where θ is the vector of parameter which characterizes the family of random variables. If $T_n(X)$ is the estimator of θ, it is said to be (weakly) *consistent* if

$$\operatorname*{plim}_{n\to\infty} T_n(X) = \theta\,.$$

In order to implement the procedure of estimation, we have chosen the following time frequencies $\Delta t = 2, 3, 4, 5, 10, 15, 20$ days. Since our least frequency available is the daily one, we cannot investigate and test the model in an intraday environment. Due to invariance under scaling of the model, it might be a very interesting field to study for operative applications. In addiction, we have $T = 1450$ days.

Unlike the previous cases, there are not pre-set R functions to estimate the parameters of the MMAR. Hence, these have been built by us.

The values the scaling functions $\tau_X(q)$ and the prefactor functions $c_X(q)$ are computed are from $q = 0.01$ to $q = 10$, with subsequent steps of 0.01. The values for $q = 0$ are provided by theoretical argumentations (see Section 3.3 and Appendix C.1), and are equal to $\tau_X(0) = -1$ and $c_X(0) = 1$ respectively.

We remind that we have a multifractal structure only if the scaling function is concave, and hence non-linear. As a matter of fact, the presence of a linear scaling function is typical of (uni)fractal stochastic processes, such as Browninan motion and fractional Brownian motion. Their scaling functions are

$$\tau_X(q) = \frac{1}{2} \cdot q - 1 \qquad \text{and} \qquad \tau_X(q) = H \cdot q - 1$$

respectively.

Figure 4.10 shows the graphical estimation of the scaling functions. As we can see, all are concave and cross the horizontal axis around $q = 2$. This fact will be crucial hereafter. But, before giving an explanation of it, we concentrate on two undeniable facts provided by visual inspection.

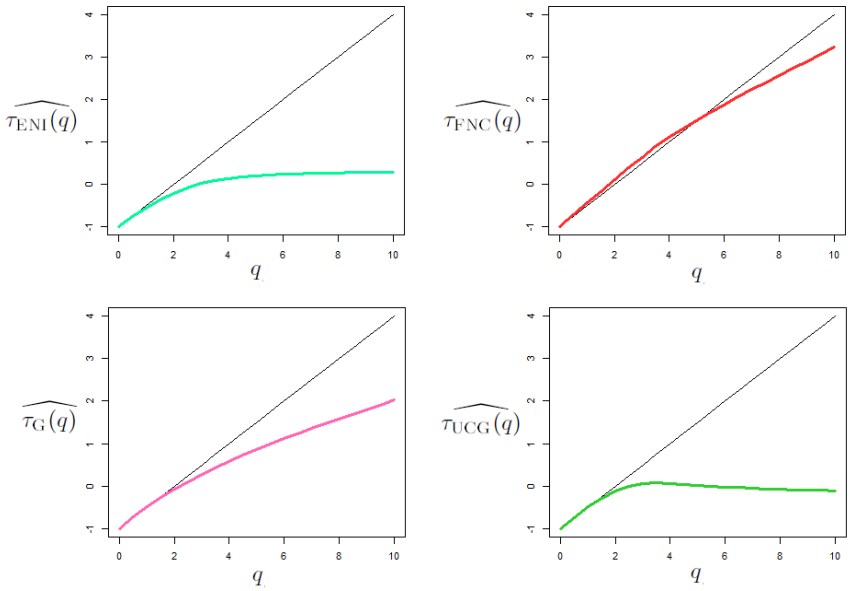


Figure 4.10: The estimated scaling functions for the four stocks.

Firstly, because of non-linearity of the estimated functions, Brownian motion scaling function (which is the straight line in the graphs) and fractional Brownian motion one are to be excluded. Thus, these models should not be employed to describe the logarithms of the prices.

The second remarkable fact is in UGC.MI's scaling function: This is the only one to have a maximum around $q = 3$, when the other ones are always increasing. Hence, $\widehat{\tau_{\mathrm{UCG}}(q)}$ starts being increasing but ends decreasing. This eventuality is typical of heaviness in the tails of the stock's distribution. So, the estimation of the second zero of the function is required. That value is very related to the q_{crit} mentioned at the beginning of this section.

We opted for not plotting the estimated functions $\widehat{\tau_X(q)}$. The reason being is that they are not so relevant for our purpose. Anyway, the codes in Appendix D provide the esteems and automatically plot those functions.

Since visual inspection is useful but not completely satisfying, we report in Table 4.4 the values taken by the four scaling functions. In addiction, it illustrates the neighbourhood of q's values for which they become null (black boxes). For UCG.MI the same information is provided for the second zero (red boxes).

In fact, as already stated, the first point where the scaling function is equal to zero needs attention. Since $\tau_\theta(1) = 0$ and, by virtue of Theorem 3.5.2, $\tau_X(q) = \tau_\theta(Hq)$, setting $q = \frac{1}{H}$ we find

$$\tau_X\left(\frac{1}{H}\right) = 0\,. \tag{4.12}$$

Table 4.4: Estimates of the scaling functions.

q	$\widehat{\tau_{\mathrm{ENI}}(q)}$	$\widehat{\tau_{\mathrm{FNC}}(q)}$	$\widehat{\tau_{\mathrm{G}}(q)}$	$\widehat{\tau_{\mathrm{UCG}}(q)}$
0	-1	-1	-1	-1
0.5	-0.762673	-0.715516	-0.726211	-0.735908
1	-0.552367	-0.436003	-0.489933	-0.493309
1.5	-0.366777	-0.161164	-0.277907	-0.282180
2	-0.209279	0.113159	-0.082944	-0.114219
2.5	-0.081959	0.383385	0.098998	-0.000316
3	0.016088	0.641744	0.269948	0.058363
3.5	0.088634	0.882877	0.430923	0.074530
4	0.140905	1.106231	0.582609	0.065926
4.5	0.178101	1.314356	0.725756	0.046138
5	0.204566	1.510658	0.861311	0.022940
5.5	0.223587	1.698188	0.990390	0.000127
6	0.237532	1.879298	1.114182	-0.020684

Continued on next page

Continued from previous page

q	$\widehat{\tau_{\mathrm{ENI}}(q)}$	$\widehat{\tau_{\mathrm{FNC}}(q)}$	$\widehat{\tau_{\mathrm{G}}(q)}$	$\widehat{\tau_{\mathrm{UCG}}(q)}$
6.5	0.248065	2.055694	1.233837	-0.038967
7	0.256352	2.228585	1.350393	-0.054715
7.5	0.263202	2.398826	1.464724	-0.068145
8	0.269173	2.567034	1.577533	-0.079556
8.5	0.274644	2.733654	1.689354	-0.089253
9	0.279867	2.899018	1.800580	-0.097519
9.5	0.285006	3.063376	1.911483	-0.104605
10	0.290160	3.226920	2.022246	-0.110723

Through it, we can easily estimate the self-affinity index $\hat{H}$. Having seen that the this value is about 2, it mean that $\hat{H} \approx \frac{1}{2}$. If this approximation is plausible[10], due to Theorem 3.5.3 the martingale hypothesis for log-prices are extendable to the MMAR. Hence, we would deal with a model with uncorrelated increments, but long memory in the q-th absolute ones.

Table 4.5 shows the estimates of the self-affinity parameter via (4.12). Hence, we found that value of q such that $\widehat{\tau_{\mathrm{ENI}}(q)} = 0$. Its reciprocal gives an estimation of $\hat{H}$ for the MMAR. The eventuality of finding different values of $\hat{H}$ as opposed to the ones estimated in Section 4.1 for the FBM models does not have to worry nor to surprise. The MMAR provides for a multifractal structure; the FBM only a (uni)fractal one. Many information "absorbed" by the Hurst parameter in the FBM context might be gathered by the scaling function (or the multifractal spectrum). They should be expected to be similar if the function which deforms time were linear, that is $\theta(t) = t$. Moreover the estimation procedure is totally different.

Table 4.5: Estimates of the self-affinity parameter.

X	q	$\hat{H}$
ENI	2.907428	0.3439466
FNC	1.793527	0.5575607
G	2.223943	0.4496517
UCG	2.501889	0.3996979

The fact of having found even values lesser than 0.5 might be explained by the bias due to inversion of logarithm and expectation.

As we have already underlined, the scaling function of UCG.MI is different from the ones of the other stocks since there is a value of q where the it becomes equal to zero. Let we call that point q_{max}. The reason being is that

[10]The fact that $\hat{H} \approx \frac{1}{2}$ does not mean that we are facing a BM. The process is still multifractal, being the scaling function non-linear.

the scaling function itself must be restricted only to those values in $[0, q_{\max})$. Beyond these values, the absolute moments of the process become infinite due to the presence of Paratian tails in the multifractal (random) measure underlying the process of the asset returns (see Appendix C.1). Since the other three function does not have that second zero, we set $q_{max} = \infty$, being so their absolute moments finite for all values of q.

Finally, let we call q_{crit} the corresponding values of the measure $\theta(t)$. Because of the link between τ_X and τ_θ, we find the values in Table 4.6.

Table 4.6: Estimates of q_{crit}.

X	q_{max}	q_{crit}
ENI	∞	∞
FNC	∞	∞
G	∞	∞
UCG	5.502899	2.199497

The presence of heavier tails in the UCG's distribution of log-returns can be detected by Figure D.1 and D.2 (in addiction, both figures illustrate that none of the distributions is Gaussian).

The scaling law in absolute moments (3.17) is useful in generating testable restrictions, but not an intuitive description of multifractality. The deepest, though probably most difficult, understanding of multifractals comes from studying local scaling properties and the multifracal spectrum (see Section 3.4 for a hint, and Appendix C.3 for an in-depth treatise).

The concept of local Hölder exponent as defined in Definition 3.12 is crucial. Hereinafter, we recall only a few aspects of it. The local Hölder exponent $\alpha(t)$ measures the local regularity of a process

$$X(t, \Delta t) \sim \Delta t^{\alpha(t)} \tag{4.13}$$

for all t, as $\Delta t \to 0$. In a standard BM and in a standard FBM, the local variation are always proportional to $\Delta t^{\frac{1}{2}}$ and to Δt^H respectively. On the other hand, multifractal processes generate variety in local regularity while filling with a *continuum* of values of $\alpha(t)$. The multifractal spectrum $f(\alpha)$ is the main tool for describing the distribution of local Hölder exponents.

Thanks to the results in Section 3.4, we know that

$$f(\alpha) = \inf_{q \in \mathbb{R}} \{\alpha \cdot q - \tau(q)\},$$

where $\tau(q)$ is the usual scaling function. Thus, due to (2.43) $f(\alpha)$, if positive, can be seen as the Hausdorff dimension of the set of points having local Hölder exponent equal to α.

Moreover, by the virtue of Theorem 3.5.6, there exists a link between the multifractal spectrum of the stochastic process for the asset returns and the one of the multifractal underlying measure

$$f_X(\alpha) = f_\theta\left(\frac{\alpha}{H}\right) .$$

Since, the two conditions

$$f_\theta(\alpha) \leq 1 \qquad \text{and} \qquad f_\theta(\alpha) \leq \alpha$$

hold true (for its proof, see Appendix C.1 and C.3), and the spectrum has a tangent to both these lines at exactly one point, thus

$$f_X(\alpha) \leq 1 \qquad \text{and} \qquad f_X(\alpha) \leq \frac{\alpha}{H} , \tag{4.14}$$

with equality at a single point. The two tangency points of f_θ are named $\alpha = \mu_\alpha$ (with the bisector) and $\alpha = \alpha_1$ (with the line $y = 1$) respectively (for a visual help, see Figure C.2).

Due to concavity of the scaling function and to the properties of Legendre transform (see Appendix C.2), the multifractal spectrum is concave as well.

All in all, regarding the multifractal spectrum of the measure, there are four values of α which require particular attention (an in-depth motivation of it can be found in Appendix C.2 and C.3):

- μ_α denotes the value of α for which $f_\theta(\mu_\alpha) = 1$. This is the maximum value of the multifractal spectrum. By strict concavity, the maximum is unique. This is also the most frequent occurring value of α, occupying a set of Lebegue measure one (or T if the sample span is not renormalized to a value of one);

- α_{min} denotes the smallest α the multifractal spectrum is defined. Being the scaling function asymptotically linear, that value is the slope of the asymptote for $q \to +\infty$;

- α_{max} denotes the largest α the multifractal spectrum is defined. Conversely, it is the slope of the asymptote for $q \to -\infty$ (this value will not be estimated, since it would require the estimation of the scaling function for negative values of q; however, the calculation of $P(q, \Delta t)$ with $q < 0$ is not appropriate for financial data).

Among these ones, the most important for investors should be α_{min}. Since (4.13) holds for $\Delta t \to 0$ (hence $0 < \Delta t < 1$), the least exponent corresponds to the most irregular instants of the log-price path, and thus the *riskiest* events for investors. The lesser its value, the most risky the asset should be considered. Since $X(t, \Delta t)$ stands for the log-price variation $\ln[S(t)] - \ln[S(t - \Delta t)]$, as Δt goes to zero we get instantaneous price variations. If α_{min} occurs, that variation might be tremendous.

Being the slope of the right asymptote, it can be easily estimated[11]. Table 4.7 reports the estimated values. As expected, the riskiest asset is the

Table 4.7: Estimates of α_{min}.

X	$\widehat{\alpha_{min}}$
ENI	0.0155046
FNC	0.3205045
G	0.2190806
UCG	-0.0060567

one with the heaviest tails. Furthermore, UCG.MI is so risky to have even a negative value of $\widehat{\alpha_{min}}$ (remind that its scaling function has a second zero $q_{max} > 1$, being so decreasing for $q \to +\infty$).

Let we consider daily data. If $\Delta t = 1$ minute $= \frac{1}{510}$ day (since Italian stock exchange open at 9 am and close at 5.30 pm) and $\alpha_{min} = -0.00605674$ occurs, prices may change about $e^{\alpha_{min} \cdot \Delta t} - 1 = 182.49\%$ only throughout one minute. We want to highlight that, during the first week of January 2012, UCG.MI plummeted up to 183.41% (yes, "rare events" happen and really matter). None of the three models of the previous section can contemplate variations of such a magnitude.

The second crucial exponent is μ_α. Since it is the most frequently occurring local Hölder exponent, the lesser its estimation, the riskiest will be the asset. On the other hand, if it is near to $\frac{1}{2}$, it means the behaviour of the price is likely to be "quasi-Brownian" (but, with the difference that Brownian motion permits the existence of only one exponent, namely $\alpha(t) = \frac{1}{2}$ for all t, when multifractal processes allow a wider spectrum of exponents among α_{min} and α_{max}).

If α_{min} (and α_{max}) is the same for both the multifractal spectrum of the measure f_θ and the multifractal spectrum of the asset reterns f_X, we are cannot consider the other two exponent of the list at the same way. Particularly, we are interested in that value such that $f_X(\alpha) = 1$. We indicate that exponent with m_α. Figure 4.11 shows them.

As expected the plots have all a maximum m_α such that $f_X(m_\alpha) = 1$. Despite what shown by Figure C.2, these spectra behave linearly for those values $\alpha > m_\alpha$. Nevertheless, taking our stand upon the theory, the functions would be expected to decrease. This is due to the fact that we estimated the scaling function $\tau(q)$ only for positive values of q. Hence the multifractal

[11] The estimation of the minimum α is coarsely given by

$$\widehat{\alpha_{min}} = \frac{\widehat{\tau(100)} - \tau(\widehat{99.999})}{100 - 99.999}.$$

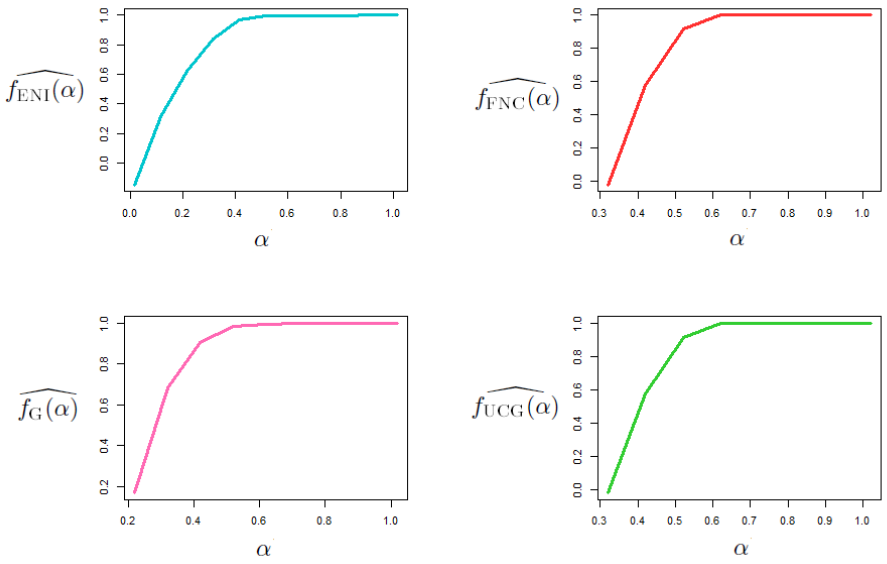


Figure 4.11: The estimated multifractal spectrum for the four stocks.

spectrum can show only its "left face". Hence, we shall consider m_α as the *first* exponent which makes the spectrum equal to one, that is

$$m_\alpha := \inf \{\alpha > \alpha_{min} : f_X(\alpha) = 1\} .$$

Having computed the whole spectrum, the estimated value $\widehat{m_\alpha}$ is trivial to find, using the inverse function of the spectrum. Moreover, since

$$f_X(\alpha) = f_\theta \left(\frac{\alpha}{H} \right)$$

holds true, we have $\mu_\alpha = \frac{m_\alpha}{H}$, where μ_α is the value such that $f_\theta(\mu_\alpha) = 1$.

Table 4.8: Estimates of m_α and μ_α.

X	$\widehat{m_\alpha}$	$\widehat{\mu_\alpha}$
ENI	0.4565046	1.3272543
FNC	0.5685045	1.0196279
G	0.5340806	1.1877651
UCG	0.6119433	1.5310146

Since the spectra have an abrupt transition at approximately α_{min} (Figure 4.11 does not show it since the values are plotted only for $\alpha > \alpha_{min}$; using codes of Appendix D, that being stated is evident), we infer that the underlying random measure is *canonical* (hence, we can address it as a *canonical*

trading time). Because of this, we are allowed to use (C.11), that is

$$f_\theta(\alpha) = 1 - \frac{(\alpha - \mu_\alpha)^2}{4 \cdot (\mu_\alpha - 1)} \,.$$

Combing with the link between the multifractal spectrum of the measure and the one of the asset returns, we get[12]

$$f_X(\alpha) = 1 - \frac{(\alpha - m_\alpha)^2}{4 \cdot H \cdot (m_\alpha - H)} \,. \tag{4.15}$$

Figure 4.12 shows that all the spectra but the one of FNC.MI are very similar to the ones provided by (4.15), which is based on theoretical assumptions. It might indicate that it is correct to suppose a canonical trading time for ENI.MI, G.MI and UCG.MI, but not for FNC.MI. As a matter of fact, looking at Figure 4.10, we can see that its behaviour is very similar to the one of a unifractal process driving log-returns. To confirm the mildness of FNC.MI, we can also see that it exhibits the highest α_{min}, and hence it can be considered as the least risky asset among those examined.

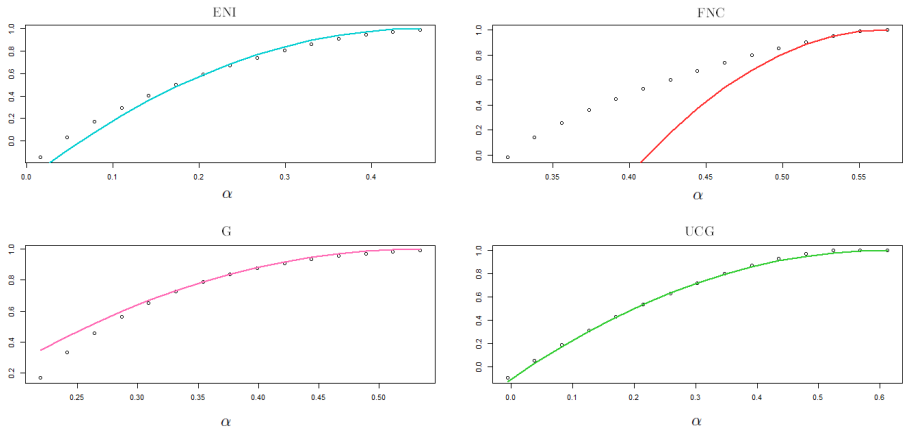


Figure 4.12: Estimated multifractal spectra vs. curve-fittings via (4.15).

Table 4.9 aims at clarifying that idea. Heuristically, the largest the range of possible $\alpha(t)$, the riskiest is the asset since the coarse Hölder exponents can take more values. It conveys a more variability of the log-returns, and hence an higher uncertainty in the magnitude of future price variations.

[12] We have just to compute

$$f_\theta\left(\frac{\alpha}{H}\right) = 1 - \frac{\left(\frac{\alpha}{H} - \mu_\alpha\right)^2}{4 \cdot (\mu_\alpha - 1)} = 1 - \frac{\frac{1}{H^2}\,(\alpha - H \cdot \mu_\alpha)^2}{4 \cdot (\mu_\alpha - 1)} \,,$$

and then use the fact that $\mu_\alpha = \frac{m_\alpha}{H}$.

Table 4.9: Estimates of the range for the admissible $\alpha(t)$.

X	$\widehat{\alpha_{min}}$	$\widehat{m_\alpha}$	$\lvert\widehat{\alpha_{min}} - \widehat{m_\alpha}\rvert$
ENI	0.0155046	0.4565046	0.441
FNC	0.3205045	0.5685045	0.248
G	0.2190806	0.5340806	0.315
UCG	-0.0060567	0.6119433	0.618
	0.0209433		0.591

Once more, UCG.MI[13] turned out to be the riskiest, while FNC.MI the one with a shorter range of admissible values.

Now we have estimated all the possible quantities required to replicate and simulate the MMAR. Since $X(t) = \ln[S(t)] - \ln(S_0) = B_H[\theta(t)]$, we find

$$S(t) = S_0 \cdot e^{B_H[\theta(t)]}, \tag{4.16}$$

as equation to generate random paths for the stocks' price. However, using the iterative version of the multifractal process instead of (4.16), the resulting dynamics can also be seen as a particular version of a *stochastic volatility* model $\sigma(t)$. Rescaling the volatility dynamics in a way to preserve a mean value equal to one (that is $\mathbb{E}[\sigma(t)] = 1$), we can write daily log-returns over unit time intervals as the product of the local volatility and the discrete version of a fractional Gaussian noise, that is

$$X(t,1) = \underbrace{\sqrt{b^k \cdot \theta_k(t)}}_{\sigma(t)} \cdot \sigma \cdot \Big[B_H(t) - B_H(t-1)\Big], \tag{4.17}$$

where

$$\theta_k(t) = \theta_k\left(\Big[t, t + b^{-k}\Big]\right) = \prod_{j=1}^{k} U_j,$$

[13]Table 4.9 shows two values of $\widehat{\alpha_{min}}$ for UCG.MI. The negative value is the one introduced firstly, as the slope of the right asymptote of $\tau(q)$. The positive one is obtained solving the equation

$$\alpha^*_{min} := \inf\{\alpha > 0 : f_X(\alpha) = 0\}.$$

The reason being is that such a procedure limits the admissible values of $\alpha(t)$, in order to be able to easily interpreter $f(\alpha)$ as the fractal dimension of the set of instant where α occurs. If we apply this procedure to the other stocks, we obtain values of very close to the one estimated through the scaling function.

Despite the fact that Large Deviation Theory (see Appendix C.3) allows for negative α, and hence negative $f(\alpha)$, Mandelbrot succeeded at introducing the concept of negative fractal dimension as "degree of emptiness of an empty set". See [9]. Negative fractal dimensions are known as *latent* α, deriving from the fact that sets of negative fractal dimension will typically not be found in a random sample. For financial data, it would be more appropriate working on *manifest* α, for which $f(\alpha) \geq 0$.

being U_j i.i.d. random variables such that[14]

$$U_j \overset{d}{\sim} \mathcal{LN}\left[-\mu_\alpha \cdot \ln(b), \sigma_\alpha^2 \cdot \ln(b)^2\right].$$

It implies that we use a finite multiplicative multifractal measure $\theta_k(t)$ (k is an integer number) in oreder to approximate the limit multifractal measure $\theta(t)$ (that is when $k \to \infty$). Furthermore, the factor b^k in (4.17) compensates for the expectation equal to $\frac{1}{b}$ of the k multipliers, σ is the same of (4.3), and H is the Hurst parameter estimated via (4.12).

Since b and k are parameters that has to be chosen *a priori* in order to do simulations[15], the only quantity that needs to be estimated is the variance of the random variables $-\log_b(U_j)$, namely σ_α^2. Using C.10 combined with C.11, we find

$$\sigma_\alpha^2 = \frac{2 \cdot (\mu_\alpha - 1)}{\ln(b)}.$$

Just to show how flexible is the MMAR as a stochastic volatility model, Figure 4.13 show six different output it can generate. The first thing that emerges is that it can achieve a large variety of different volatility structures. Secondly, it display strong volatility clustering of many sorts. However, despite GARCH moldels, the MMAR is invariant under scaling, and hence applicable at all the available frequencies.

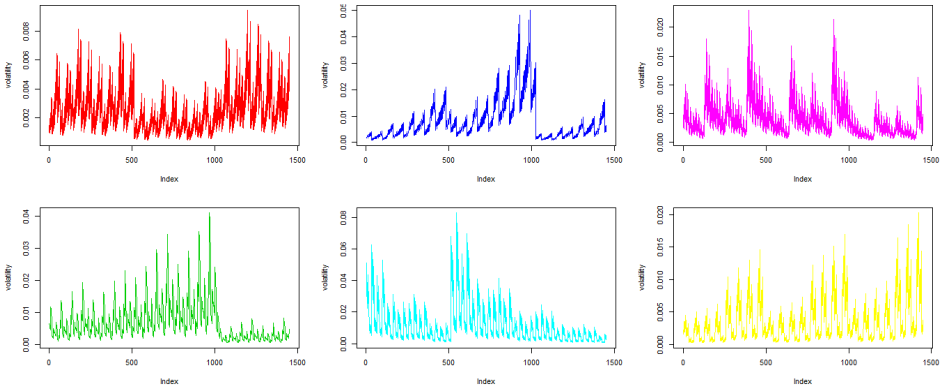


Figure 4.13: Feasible volatility's outputs.

Having estimated all the required parameters with the codes of Appendix D, we are able to replicate the past records of prices and log-returns for all the four assets.

Figure 4.14 finally shows clearly what been stated throughout all this work. While BM, FBM or LSM are not able to catch simultaneously all

[14] See Section 2.5 and Appendix C.3.

[15] Herein, we will use $b = 2$ and $k = 11$ because the total length of the time series is $T = 1450$ days.

the peculiarities of financial data – and hence we cannot expect they replicate them in a correct way – the MMAR seems to provide a more faithful reproduction of the reality.

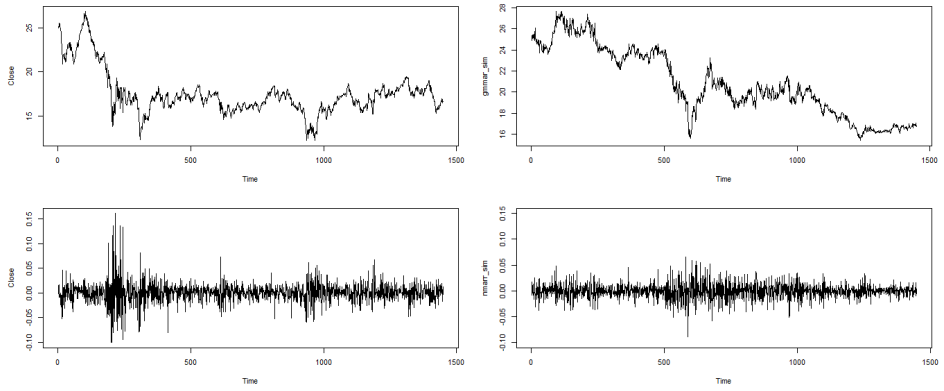


Figure 4.14: Real ENI's data on the left side; simulations to the right.

On the other hand, we are totally aware of the fact that not even the MMAR is perfect. However, among all the exhibited alternatives, it is doubtless the one that best fit actual financial data. As a matter of fact, unlike Figures 4.1, 4.4, and 4.7, Figure 4.14 shows variations of the same magnitude of real log-returns.

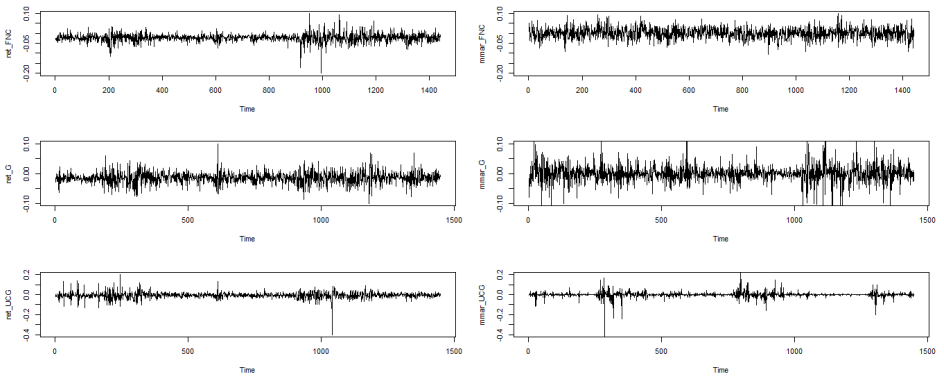


Figure 4.15: Real data on the left side; simulations to the right.

By visual inspection, we can also assert that the simulated data could exhibit an higher volatility that the real ones. This might be due to the fact that we opted for using log-normal random multipliers U_j such that $U_j \geq 0$, when for ENI.MI, FNC.MI, and G.MI it would be more appropriate to use random multipliers such that $0 \leq U_j \leq 1$ (since the relative scaling functions are always increasing). Of course, if that eventuality were verified, it would imply the risk, seen as the volatility of log-returns, to have been slightly

overestimated (when the previous fractal models tended to underestimate it).

Figure 4.16, compared with Figures 4.3, 4.6, and 4.9, shows very well how this multifractal process can take into account all the peculiarities of financial data, namely scale-invariance, persistency, heavy tails, and the existence of many different local behaviours. Hence, the common excuse of the unpredictability of *structural breaks* and the presence of *outliers* fall to pieces here. They are an unavoidable part of financial time series, and the MMAR – with its variety of volatility's structures – is able to work that issue out.

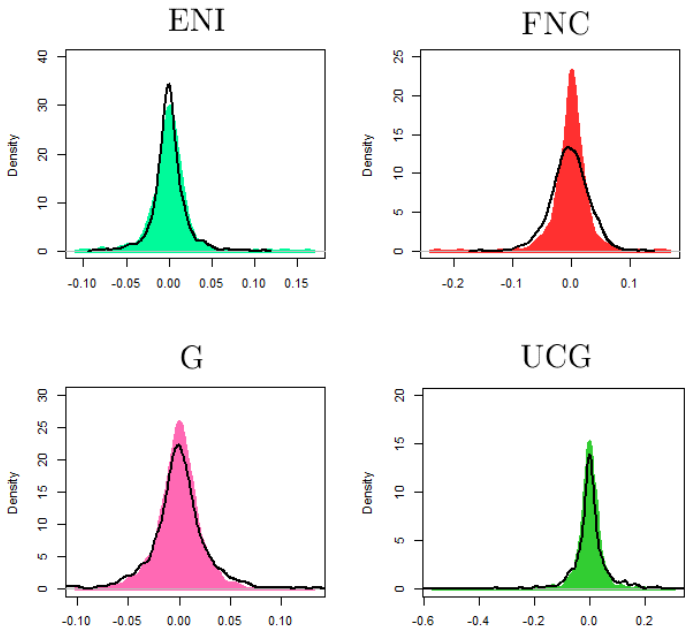


Figure 4.16: Densities of real data vs. densities generated by the MMAR.

As throughout the whole section, we see that the worst represented asset is FNC.MI. As a matter of fact, perhaps the MMAR overestimates the risks generating a path which has got heavier tails than necessary. Actually, looking at Figure 4.3, we can state the BM seems to fit it pretty well. However, we want to stress again that this sort of bias might be given an overestimation of $\hat{H}$ (Figure 4.10 shows that its scaling function is very close to the one of a BM, that is $H = \frac{1}{2}$). Moreover, a "milder" choice of the masses U_j might be more appropriate (but, at the moment, there are not known expressions for generating U_j according to the multifractal spectrum $f(\alpha)$).

In addiction, for practical purposes overestimating the risk might be less dangerous than underestimating, even though such an eventuality might become very onerous for those financial institutions required to allocate amounts of money to cover the *market risks*.

The different quantifications of the risks will be crucial in the final section when applied to option pricing. That is nothing more than giving the *fair price of risk* for the underlying asset.

4.3 Option pricing outcomes

In this last section, we will apply the four different models to a real option pricing issue. In particular, we will compare the different prices of a American call written on the same four asset taken into account in this chapter. The choice of an American call, rather than an European one, is due to the fact that Italian stock exchange quotes only this type of call. In fact, our results based on BM, FBM, LSM, and MMAR will be eventually compared to the real market quotation. In some way, this comparison want to inspect what kind of risk perception the markets agent experience (and hence what kind of pricing models they might adopt).

However, given the absence of arbitrage opportunities, for American call options the early exercise is never advantageous, unless the underlying asset provides for dividends payment. Thus, if no dividends are expected, the price of an American call option is the same of an European one with same strike price and expiration date. That is

$$C(t) = c(t)$$

where $C(t)$ is used to indicate the price of the American option at time $t \in [t_0, T]$ and the lower $c(t)$ the European one at the same time.

Hereinafter, we will assume that the four stocks do not pay dividends (this assumption is also plausible since Italian stocks exhibits an almost-zero dividend yield).

The employed technique to estimate $C(t)$ will be the Monte Carlo simulation. Within this environment, the procedure to generate the paths of the underlying asset's price $S(t)$ is crucial. Moreover, random paths have to be produced depending on the eventuality to be allowed to operate in a *risk-neutral economy*. The effectiveness of that hypothesis, which should be always proved[16], is sure for an European call when prices follows a geometric Brownian motion. As a matter of fact, it is always possible to build an hedging portfolio that replicates the call's payoff exactly. For the other stochastic processes, we will postulate to be able to operate in the same way.

If it is right to suppose that investors are risk-neutral, calculations involved in the evaluation of derivatives oversimplify. As a matter of fact, the expected rate of return of each asset is the same of the risk-free asset. Furthermore, in a risk-neutral risk market the value at time $t \in [t_0, T)$ of a

[16]Such a proof consists in showing that there exists an *equivalent martingale measure* for the stochastic process used to describe prices' evolution through time (see Girsanov theorem).

Table 4.10: Analytical expressions of the four models.

Model	Equation
geometric Brownian motion	$S(t) = S_0 \cdot e^{\left(r - \frac{\sigma^2}{2}\right) \cdot t + \sigma \cdot B(t)}$
geometric fractional Brownian motion	$S(t) = S_0 \cdot e^{\left(r - \frac{\sigma^2}{2}\right) \cdot t + \sigma \cdot B_H(t)}$
geometric Lévy-stable motion	$S(t) = S_0 \cdot e^{\left(r - \gamma^2\right) \cdot t + \sqrt{2} \cdot \gamma \cdot M(t)}$
geometric MMAR	$S(t) = S_0 \cdot e^{B_H[\theta(t)]}$

random cash flow available at time T can be easily estimated discounting the expected value of the cash flow itself using as discounting rate the risk-free rate.

Hence, if risk-neutral hypotheses hold, the value of a (European or American with no dividends) call option at time t is given by

$$C(t) = e^{r \cdot (T-t)} \cdot \mathbb{E}\left[C(T)\right] , \tag{4.18}$$

where r is the risk-free log-return rate, and

$$C(T) = \max\left\{0, S(T) - K\right\} \tag{4.19}$$

is the call payoff at expiration date (of course, K is the strike price). In our analysis, we have estimated the risk-free rate using EURIBOR rates[17]. Given the estimate $\hat{r}$, this means that $\hat{\mu}$ of Section 4.1 has to be substituted by that value[18].

Below, the adopted Monte Carlo procedure is described. For each price evolution structure listed in Table 4.10, we are able to generate M different paths of the underlying asset price $S(t)$. At expiration date T we compute the payoffs of the call for all the different $S_i(t)$, with $i = 1, 2, \ldots, M$. Then the arithmetic mean is used as estimator of the expected value, allowing us to get $\widehat{\mathbb{E}\left[C(T)\right]}$, that is

$$\widehat{\mathbb{E}\left[C(T)\right]} = \frac{1}{M} \cdot \sum_{i=1}^{M} \max\left\{0, S_i(T) - K\right\} . \tag{4.20}$$

Finally, the estimated price of the call option at time t is obtained via (4.18).

[17] Date are taken from `http://www.itistimed.com/gadgets/euribor-history.php?Y=2013`.

[18] We have to underline that such a substitution cannot be adopted with the MMAR since the model does not have a manifest drift term. Anyway, payoff at time T will be discounted using $\hat{r}$ even for that model.

Table 4.11: Call price and stock price at time $t_0 = 20/09/2013$; strike price for expiration date $T = 15/11/2013$.

Stock	C_0	S_0	K
ENI	0.2530	17.95	18.00
FNC	0.1030	4.52	4.80
G	0.3210	15.00	15.50
UCG	0.2015	4.86	5.00

Market quotations of the calls' prices refer to $t_0 = 20/09/2013$ with expiration date $T = 15/11/2013$ (45 working days). Since a volatility parameter has to be estimated for those models that consider a constant volatility structure (namely σ for BM and FBM, γ for LSM), in order to get more representative and effective estimations of them we will follow the most common approach which requires to use a time series with a length very close to the number of days in the interval $[t_0, T]$. The estimation of the risk-free rate of return has been computed using the quoted rate at t_0 for that maturity closest to the expiration day of the option[19]. Particularly, the daily rate is equal to

$$\hat{r} = 0.00003\,.$$

As opposed to those models, since MMAR generates a stochastic volatility structure, we opt for using the entire time series starting from 01/01/2008 (moreover, the estimations of the scaling functions and the multifractal spectra require a large amount of data).

Table 4.11 contains market prices of the calls at time 20/09/2013 for the relative strike prices. In addiction, we reported the closing price of the stock in the same day.

In implementing Monte Carlo simulations, we decided to set $M = 500000$ which means that the expected call's price $\mathbb{E}\widehat{[C(T)]}$ is computed using five hundred thousands different paths (and hence five hundred thousands different final prices $S(T)$). This large number of paths ought to assure the convergence to the real value $\mathbb{E}[C(T)]$.

Following [9], in order to simulate the MMAR more accurately, we opted for setting the memory parameter equal to $H = \frac{1}{2}$ in equation (4.17). This choice is motivated by three different considerations:

[19] On 20/09/2013 the 1-month and 2-month EURIBOR were equal to 0.128% and 0.174% respectively. Since the expiration day is 45 days after t_0, with a rough linear interpolation, we get 45-days EURIBOR equal to $0.5 \cdot (0.174 + 0.128)\% = 0,15\%$. The daily value is obtained via the following financial equivalence

$$\hat{r} = \sqrt[45]{1 + r(t, 45)} - 1\,.$$

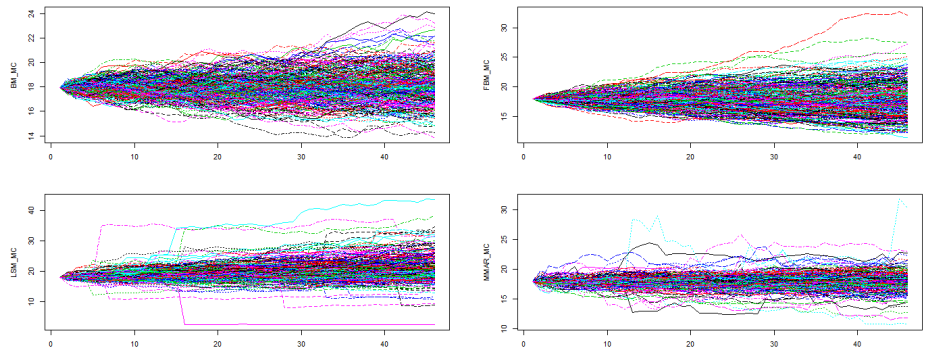


Figure 4.17: Simulated paths for ENI.MI
(only the first 300 ones are plotted).

- The estimations of H in Section 4.2 are all close to 0.5, and hence, despite the fact that a statistical testing procedure does not exist yet, we think we could not reject the null hypothesis $H = \frac{1}{2}$ in favor of the alternative hypothesis $H \neq \frac{1}{2}$ (in addiction, see footnote 8 of Section 4.2);

- The fact of setting $H = \frac{1}{2}$ allows us to extend the martingale hypothesis for log-prices (and hence the non-arbitrage principles of asset pricing should hold true). We highlight that in this model martingale hypotheses and long memory in squared returns coexist via a multifractal trading time;

- The MMAR seems to fit real data reasonably well even under this restriction.

For the other models, no restriction under the estimated parameters was done[20]. Outputs using FBM as driving process have to be considered very carefully, since the presence of H may cause systematic arbitrage strategies (prices, being dependent, can be foreseen).

Monte Carlo simulations have provided the estimations listed in Table 4.12, although with different level of computational complexity[21].

[20] Only in the LSM case a non-parametric restriction has been necessary. Because of its extreme wildness (that is, it generates tails too heavy), we noted the process reached "unacceptable" prices, in the sense that upper variations might be too high leading to prices of the magnitude of 10^3. Of course, the procedure provided call's prices of the same magnitude.

In order to avoid such a drawback, we added the following rough restriction

$$S^*(T) = \min\{S(T), 4 \cdot S_0\} ,$$

which allows for both heavy tails and reasonable prices.

[21] Of course, the model that requires the largest amount of data and calculations is the

Table 4.12: Call prices at time $t_0 = 20/09/2013$ for all the models.

	ENI	FNC	G	UCG
Black-Scholes	0.2171	0.1135	0.3017	0.2079
BM	0.2174	0.1143	0.3025	0.2083
FBM	0.3289	0.3210	0.5723	0.3658
LSM	1.0490	0.0848	0.0648	0.0150
MMAR	0.8608	0.2350	0.5798	0.5816
Market	**0.2530**	**0.1030**	**0.3210**	**0.2015**

As we can see, FBM and LSM generate very high/low prices of $\widehat{C_0}$. This can be ascribed to two different reasons: For FBM, the estimations[22] of H might cause an inconsistence in using such a model for fair option pricing (in fact, the martingale property fails); in LSM case, the estimations of the parameters of the stable distributions (α and γ in particular) cause a too high degree of "wildness"[23], leading to extremely high/low levels of prices. These two facts generate an over/underestimated risk, and hence very high/low price for the American call. For these reasons, we might conclude that both FBM and LSM are not so appropriate for such a purpose.

On the other hand, the MMAR seems to perform in a very efficient way. Of course, the obtained prices are higher than the one provided using a BM for the log-prices, but it can combine different sources of risks (fat tails and dependence in the absolute moments of the distribution) without apparent biases or overestimation. The differences among them could very large (for ENI.MI and UCG.MI, the price got via MMAR is thrice higher than the one provided by BM). In any case, now more than ever, a *correct pricing of the risks* embodied by derivatives is a crucial concern. Even light discrepancies in pricing are able to cause tremendous consequences over both the agent operating in financial markets and the real economy as well.

Of course, since usually market operators follow mainstream techniques, the prices of the derivative are always lower than the one obtained through the MMAR. This might indicate, especially when market prices are lower than the ones obtained using a BM, that operators *underestimate* the assets' risk in a very critical fashion.

Finally, we have to underline that the relation $C(t) = c(t)$ can be considered just an approximation for ENI.MI. Since on 26/09/2013 a dividend was expected, $C(t) > c(t)$ holds true.

MMAR.

[22] On the considered periods, we have got the following estimation for H: 0.5998, 0.6583, 0.6075, 0.6052; and for α: 1.879, 1.348, 1.745, 1.488.

[23] In fact, if α is lower than 1.8, the tails become very heavy. If the price approaches to zero, it is almost impossible to raise.

Conclusions

Throughout this work we have introduced several stochastic processes and different mathematical tools for describing the (mis)behaviour of financial prices.

In Chapter 1, firstly we handled **Brownian motion** which is the most widespread process, employed by the financial community for data modelling and forecasting. It owns several suitable properties, such as self-affinity and stationary increments, but the volatility it takes into account (embodied by the constant diffusion parameter, which is related to the variance of the log-returns) is too mild for an effective characterization of financial reality. Secondly, we introduced two more sophisticated processes such that BM is a particular case. They are the **fractional Brownian motion** and the **Lévy-stable motion**. The former conceives a long-range dependence parameter H, which works also as self-affinity parameter, thus introducing memory into log-returns. The latter, which is a particular Lévy process based on stable distributions $\mathcal{S}(\alpha, \beta, \gamma, \delta)$, takes into account heavy-tails (via the shape parameter α) and asymmetry (via the skewness parameter β). Even in this case, BM can be traced as a subcase of the large variety of processes that stable distributions are able to generate.

In Chapter 2, we introduced the main mathematical tools of **Fractal Geometry**. Despite this topic might seem unrelated to financial modelling or too theoretical for practical applications, it turns out to be crucial for a deep understanding of financial charts. Far from being self-similar (which implies the application of an isotropic transformation), they are statistically self-affine. As a matter of fact, Hausdorff dimension of those graphs is always a number taking values in the open interval the open interval $(1, 2)$. Moreover, we proved that both BM, FBM and LSM are all fractals with Hausdorff dimensions equal to $\frac{3}{2}$, $2 - H$, and $\max\left\{1, 2 - \frac{1}{\alpha}\right\}$ respectively. Finally, we gave a very theoretical dissertation about **multifractal measures**, which are deeply involved into the last model for financial modelling which this work analyzes. Crucial to underline, if fractality might be a feature of *sets*, multifractality is a property regarding *measures*. Multifractal measures are hence able to "create" different local fractals according to different local Hölder exponents. Such measures were introduced since the last model for describing financial returns is based on them.

In fact, in Chapter 3 we deeply analyzed the **multifractal model of asset return** by Calvet, Fischer and Mandelbrot. It is based on the assumption that financial data come up with local fractal structure, hence suggesting that a multifractal measure is involved into the process of price making. Such a measure is assumed to be a self-similar random (multifractal) measure deforming trading time. In other words, time is seen as a random variable whose distribution is based on a multiplicative random cascade. Many of these notions find extensive explanations in Appendix C. The fact of dealing with such a measure implies the existence of a multifractal process both in trading time and in log-returns of financial assets (while the former moldels could be addresses as *unifractal* processes). Such a process is able to meet both long-range dependence, heavy-tails, and of course self-affinity. Moreover, since memory is given by both the Hurst parameter H and the fact that absolute q-th moments of price increments contain information, under suitable restrictions, the model can both have memory and uncorrelation in price increments themselves (thus allowing the possibility for returns to be white). Furthermore, via the scaling function $\tau(q)$ and the multifractal spectrum $f(\alpha)$, we can find all the admissible local exponents (and hence all the possible behaviours of asset's returns), being so able to associate to them the relative probability of occurrence.

In Chapter 4, we decided to find the *best model* for describing four Italian stocks' past records. In order to to that, we took for each asset the corresponding time series (starting from 01/01/2008 and ending at 30/07/2013), and then estimated al the required quantities to replicate the past records. The goodness of fitting was judged by both visual inspection of real log-return vs the generated ones, and via the histogram of real data vs densities obtained by the estimated values. Our conclusions are that, if BM and FBM underestimate the weight of the information contained inside the tails, LSM produces too wild paths overestimating the occurring of rare events. On the other hand, the MMAR seems to fit very well for most of the stocks. The only asset which seems to be fitted worst is Finmeccanica's stock since, during the period of analysis, it behaved not so wildly as the other ones. As a matter of fact, having chosen to implement the MMAR via a canonical cascade (which allow for the highest degree of randomness), it maybe have caused an overestimation of the risk embodied by the asset.

The second part of the chapter was devoted to a practical application: *Option pricing*. Using the same stocks, we compared the prices of the corresponding American calls quoted by the market with those obtained via Monte Carlo simulations. As expected, while BM (and hence Black-Scholes formula) underestimate the risk with respect to the MMAR, both FBM and LSM ended up with being unappropriate (for different reasons) for such a purpose. Hence the only possible alternatives among those analyzed were the Brownian motion and the multifractal model of asset returns. Unexpectedly, the market operators pricing the call options sometimes align themselves to

BM's outputs, but more frequently they quote lower prices than the ones provided by the Black-Scholes formula. It means that their perception of risk is very low, with potentially highly dangerous effects over both their estates and the overall economy.

Furthermore, even though we did not explicitly compare the stochastic volatility produced by the MMAR with the one generate by a GARCH model, we can underline the following well-know differences. While MMAR's volatility is produced by a multiplicative random cascade which is based on data at different frequencies, if we try to estimate the parameter of a GARCH model using daily rather than weekly data, we surely find different estimations of the same parameters (if not even the disappearance of the GARCH effects). For the same reason, MMAR outputs are **invariant under scaling** when GARCH's ones can be only used at the same frequency they were estimated by. In addiction, it is quite frequent for a GARCH(1,1) not to be able to fit the data effectively. That is the reason why many extensions of original GARCH models were suggested by many authors: TGARCH for taking into account the asymmetry of data, EGARCH for those cases where the GARCH's estimations brake the restrictions of positivity, IGARCH for modelling a non-stationary variance, and so on and so forth... All remarkable efforts, but no one able to catch all the peculiarities of actual data. Eventually, GARCH models consider volatility as a bounded ARMA model of the *second moment* of the return's distribution. MMAR volatility is based on an unbounded estimation procedure which uses untill the absolute *q-th moment* of the same distribution.

Risk is the milestone of Finance and Financial Economics. We have shown over and over that prices do not vary mildly and continuously, but they are subjected to wild fluctuations at all time scales. Volatility – far from being a static entity to be ignored or easily compensated for – is at the very heart of what goes on in financial markets. In the past, nearly everyone embraced the Modern Portfolio Theory because of the absence of strong alternatives. But one need no longer accept it at face value. Modern Portfolio Theory inherited a large number of techniques that statisticians designed to deal with mild Gaussian variability. The challenge was to adapt them to the context of financial prices. Now, the same effort should be requested at an higher liver, in order to reformulate new financial theories according to those evidences stated throughout this work.

A possible and really interesting idea was conceived by Benoît Mandelbrot in [24] using some notions of Actuarial Mathematics. An enlightening analogy and powerful guidance for the future is provided by a distinction between different levels of insurance that relates to the distinction between mild and wild *state of randomness*. Most life, car, or homeowner risks are mild. Very much like the coin-tossing model, they fall within a narrow range

and are mutually independent. Even when a portfolio of insurance contracts is small, the properties of diversification (due tue the Law of Large Numbers and related results) can be trusted to create a risk of ruin that is sufficiently small to be profitable even for an insurance company with limited reserves.

In order not to risk and to insure occasional higher risk, the insurer of mild risk will seek reinsurance. One reason why reinsurance need not be expensive is because there is also an insurer of last resort. Indeed, insurers know well that, next to *diversifiable risks*, many other risks seeking to be insured are wild, (namely *non-diversifiable risks*) very like in the MMAR. They involve greatly magnified equivalent of the notorious "ten sigma" price change. Ordinary diversification would be defeated by such risks – even if the number of cases had sufficed for a law of large numbers to be trusted to apply. Inserting the odds of those wild risks, in the usual calculations of premiums – that is nothing more than the *pricing of the risk* – would imply reserves that are far too large to be reasonable. However, such risks are insurable because they are immediately shared with *reinsurers*.

The key fact is that insurers cannot survive by only considering the "fair weather" 95% of the claims, which would have easily been diversified. Not only the 5% of large "foul weather" claims cannot be ignored, but their odds are non-negligible and are the essence of those professions: They are an essential input of planned and carefully priced reinsurance. Reinsurance should be a powerful example to give a more in-depth understanding of how to rationalize the handling of "foul weather" in Finance.

Moreover, as a result of globalization, the relevance of the preceding comments to insurance is likely to increase: The larger the markets, the greater the attention demanded by the potentially disastrous effects of financial downturns.

At the end of this dissertation we can certainly state that neither BM nor the other alternatives FBM and LSM give an effective description of all the properties owned by real data. On the other hand, the MMAR seem to be very promising by its capability to adapt itself to different behaviours and to take into account all the required peculiarities of financial prices. Certainly, it is far from being perfect, and hence will require several improvements and extensions in order to fit also milder data. For the future, particular efforts should be addressed to *estimation techniques* for a more accurate and unbiased estimations of the scaling function $\tau(q)$. On the other hand, alternative *closed expressions* for the multifractal spectrum $f(\alpha)$ need to be found in order to connect the the estimated α to the distribution of the multifractal measure even for micro-canonical cascades.

Surely, the duo multifractals-financial data is very promising since with only few (even if non-standard) entities, it can achieve an actual degree of randomness. Moreover, with the same quantities it provides very different structures of stochastic volatility, being only partially linked to the second-

order moment of the distribution of the log-returns, but taking into account the information of higher-order absolute moments as well.

No overall mathematical techniques come to forecast a price drop or rise on a specific day on the basis of past records. On this account, the MMAR does not do any better. But it provide *realistic estimates* of the probability of what the market might do in the future and allow us to prepare for inevitable, potentially dramatic, changes. This new modelling techniques are designed to give a valid alternative to cast a light of of order into the seemingly impenetrable thicker of financial markets. They also recognize the mariner's warning that, as recent events demonstrate, deserve to be taken into account:

On even the calmest sea, a gale may be just over the horizon.

Mandelbrot, B.

Acknowledgements

After this very long "journey", I finally find myself able to thank those who walked with me to the end.

First, heartfelt thanks go to all close to me, who gave me their support each and every day. You know who you are, and you know the myriad ways you have helped and the magnitude of that help.

Three people in particular merit individual praise for their variety of contributions, and I will thank them in the order of their appearance on my life-stage.

The first is my high school mathematics teacher, Stefano Ravasi. He had a huge role in my decision to study finance, from his first lesson on natural numbers to the last on improper integrals. He taught me more than mathematics. He taught me how to distinguish between math and *fine math*, instilling passion for that subject - the Subject[24]. He made me aware of the importance and centrality of mathematics in both natural and social sciences. I was very lucky to have met him - and to have built a long-lasting friendship.

I found the same passion for teaching, and dediction to his students in my second mentor. Professor Paolo Pellizzari - the supervisor for my Bachelor's thesis. He proved to be one of the most kind and generous people I have ever met. He taught me how to write at an "higher" level, in a way that would be appreciated by an audience of experts. He was instrumental in bringing this work to completion, and without that help Chapter 4 would have not been possible.

Finally, Professor Marco Tolotti was vital to the completion of this opus, and in my personal growth. He was the primary supervisor and initial "victim" of my first drafts - those difficult attempts to truly understand fractals sets and multifractality. He made the theories not just a list of properties or theorems, but that fine math I hope to have been able to write. He taught me that precision and rigor must never be sacrificed to develop simpler, inaccurate expositions on any subject. This requires every assertion to be

[24] "*Die Mathematik ist die Königin der Wissenschaften*" Carl Friedrich Gauss.

proved and nothing to be taken for granted. Professor Tolotti demonstrated the patience of a saint during our long - very long!- encounters.

Others to recognize and thank include Yves Guivarc'h, Professor Emeritus of Ergodic Theory at the Institut de Recherche Mathématiques de Rennes, University of Rennes 1, and Thomas Lux, professor of Monetary Economics and International Finance at University of Kiel. Professor Guivarc'h was kind enough to introduce me to the hidden-gem of an article ([21]). Professor Lux explained the use of MMAR to generate the stochastic volatility of (4.17).

Finally, I should also thank Life and myself. Life, for what it forced me to face – tough situations at times, and their resolutions. Myself for the way I reacted, always searching for the upside which makes the difference between "living" and "getting by".

I forge ahead.

Appendix A

Appendix

A.1 Elements of probability theory

In this appendix we introduce some of the most important elements used in Probability Theory. For for an extensive discussion of the matter, see [18] and [19].

The set of all possible outcomes of a experiment is called *sample space*, denoted by Ω. Questions concerning the outcome of an experiment can always be phrased in terms of subsets of Ω. Nevertheless there is a collection $\mathcal{F}$ of subset of Ω particularly important, whose element are called events; it results to be a *σ-algebra* on Ω [1]. Next we associate probability with the events A_i of $\mathcal{F}$, with $\mathbf{P}(A_i)$ thought of as the likelihood that the event A_i occurs. We call $\mathbf{P}$ a *probability measure* on Ω if it assigns a number $\mathbf{P}(A_i)$ to each A_i in $\mathcal{F}$, such that:

$$0 \leq \mathbf{P}(A_i) \leq 1 \quad \forall A_i \in \mathcal{F}$$

$$\mathbf{P}(\Omega) = 1$$

$$\mathbf{P}\left(\bigcup_{i \in \mathbb{N}} A_i\right) = \sum_{i \in \mathbb{N}} \mathbf{P}(A_i)$$

if $A_1, A_2, ..., A_i, ...$ are disjoint event in $\mathcal{F}$.

[1] Let Ω be some set and $\mathcal{P}(\Omega)$ represents its power set. Then a subset $\mathcal{F} \subseteq \mathcal{P}(\Omega)$ is called *algebra* if it satisfies the following three properties:

- $\Omega \in \mathcal{F}$;
- $\neg A \in \mathcal{F} \quad \forall A \in \mathcal{F}$;
- $A \cup B \in \mathcal{F} \quad \forall A, B \in \mathcal{F}$.

In addiction, we say that $\mathcal{F}$ is a *σ-algebra* if it satisfies the following property as well:

- $\bigcup_i A_i \in \mathcal{F} \quad \forall \{A_i\}_{i \in \mathbb{N}} \in \mathcal{F}$.

We call a triple $(\Omega, \mathcal{F}, \mathbf{P})$ a *probability space* if $\mathcal{F}$ is an event space of subset of Ω and $\mathbf{P}$ is a probability measure define on the sets of $\mathcal{F}$.

Given a finite number of events $A_1, A_2, ..., A_n \in \mathcal{F}$, they are told *mutually independent* if and only if:

$$\mathbf{P}\left(\bigcap_{i=1}^{n} A_i\right) = \prod_{i=1}^{n} \mathbf{P}(A_i)\,.$$

In other words, these are events whose occurrence or non-occurrence is not in any way influenced by the occurrence or non-occurrence of the other ones, leading no consequences on the measure of their probability.

Furthermore, we say that a particular property related to few $\omega \in \Omega$ is valid *almost surely* (or shortly a.s.) on Ω if it is true except for a $A' \in \mathcal{F}$ such that $\mathbf{P}(A') = 0$, (where A' is often called zero-probability event).

Let $\mathcal{A}$ be a family of subset of Ω. The smallest σ-algebra which contains $\mathcal{A}$ is called σ-algebra generated by $\mathcal{A}$ and is denoted by $\sigma(\mathcal{A})$. Given a metric space E, we indicate with $\mathcal{B}(E)$ the smallest σ-algebra which contains all the open sets of E (and it is called Borel σ-algebra).

A measurable function[2] $X : \Omega \to \mathbb{R}^n$ is called *random variable*. As a matter of fact, it is requested that, for all $B \in \mathcal{B}(\mathbb{R}^n)$, $X^{-1}(B) \in \mathcal{F}$. This is usually also expressed by saying that X is $\mathcal{F}$–measurable. If $n = 1$, we are dealing with a one-dimensional real random variable.

Let $X : \Omega \to \mathbb{R}^n$ be a function. The smallest σ-algebra that makes X measurable is the σ-algebra generated by X, that is $\sigma(X)$. It is clear that $\sigma(X) = \left\{X^{-1}(B)\right\}_{B \in \mathcal{B}(\mathbb{R}^n)}$. In addiction, if $B \in \mathcal{B}(\mathbb{R}^n)$, it is commonly used to write $(X \in B)$ to indicate the set (or the event) $X^{-1}(B) = \{\omega \in \Omega : X(\Omega) \in B\}$. Furthermore, we define:

$$\mathbf{P}(X \in B) := \mathbf{P}(\{\omega \in \Omega : X(\Omega) \in B\})\,.$$

The distribution (or law) of X is the probability measure π_X on $\mathcal{B}(\mathbb{R}^n)$ given by:

$$\pi_X(B) = \mathbf{P}(X \in B), \quad B \in \mathcal{B}(\mathbb{R}^n)\,.$$

If X is Lebesgue integrable on $(\Omega, \mathcal{F})$, that is to say that the Lebesgue integral $\int_\Omega |X(\omega)|\,\mathbf{P}(\mathrm{d}\omega)$ results finite, we say that X is an *integrable random variable*. In this case, it is defined

$$\mathbb{E}(X) := \int_\Omega X(\omega)\,\mathbf{P}(\mathrm{d}\omega) \tag{A.1}$$

[2]Let (X, Σ) and (Y, Ξ) be measurable spaces, meaning that X and Y are sets equipped with respective σ-algebras Σ and Ξ. A function $f : X \to Y$ is said to be measurable if $f^{-1}(\xi) \in \Sigma$ for every $\xi \in \Xi$.

which is said *expected value* of X. Its calculation can always be traced back to the integral computation on $\mathbb{R}^n$ due to a well-know calculus result (Change of Variables Theorem). Let X and Y be two real random variable, such that $\mathbb{E}(X \cdot Y) < \infty$. The *covariance* between X and Y is given by:

$$\mathbb{C}\mathrm{ov}(X, Y) := \mathbb{E}\big\{[X - \mathbb{E}(X)] \cdot [Y - \mathbb{E}(Y)]\big\} \tag{A.2}$$

which can be written as $\mathbb{C}\mathrm{ov}(X) = \mathbb{E}(X \cdot Y) - \mathbb{E}(X) \cdot \mathbb{E}(Y)$ as well. If we calculate the covariance between a variable X and itself, we obtain the *variance* of it, that is:

$$\mathbb{V}(X) = \mathbb{C}\mathrm{ov}(X, X) = \mathbb{E}\left\{[X - \mathbb{E}(X)]^2\right\} . \tag{A.3}$$

In the same way, it can be show that the decomposition $\mathbb{V}(X) = \mathbb{E}(X^2) - [\mathbb{E}(X)]^2$ is still valid.

Let we consider a one-dimensional real random variable X and the event $A = \{X \leq x\} \in \mathcal{F}$. The probability measure on A is defined by:

$$\mathbf{P}(A) = \int_A f_X(x)\,\mathrm{d}x = \int_{-\infty}^{x} f_X(h)\,\mathrm{d}h$$

where $f_X(x)$ is called *probability density function* (shortly p.d.f.) of X (note that this does not mean that a closed expression for $f_X(x)$) exists. Furthermore, the probability measure can be associate to a function:

$$\mathbf{P}(A) = \Pr\{X \leq x\} = F_X(x) \tag{A.4}$$

whose name is *cumulative distribution function* (briefly c.d.f) of X. To say that X has c.d.f. $F_X(x)$, we will write $X \overset{d}{\sim} F_X(x)$. For these reasons, the expected value of a real random variable can be also written as $(\Omega \subseteq \mathbb{R}, X(\omega) \in \mathbb{R})$:

$$\mathbb{E}(X) = \int_{-\infty}^{+\infty} x\,\mathrm{d}F_X(x) = \int_{-\infty}^{+\infty} x\,F'_X(x)\,\mathrm{d}x = \int_{-\infty}^{+\infty} x\,f_X(x)\,\mathrm{d}x\,.$$

With this expression for the expected value, it is also easier to introduce the k-th moment of the random variable X. It is defined as:

$$\mathbb{E}\big(X^k\big) = \int_{-\infty}^{+\infty} x^k\,f_X(x)\,\mathrm{d}x\,.$$

A general way to *fully* describe a random variable is to study all of its moments. Fortunately, most random variables require only to know moments of the first and second order.

Eventually, given the same one-dimensional random variable X, let we consider another very important function called *characteristic function*. This is a function $\varphi : \mathbb{R} \to \mathbb{C}$ defined as:

$$\varphi_X(u) := \mathbb{E}\left(e^{\imath u X}\right) = \int_{-\infty}^{+\infty} e^{\imath u x} f_X(x)\,\mathrm{d}x \tag{A.5}$$

where $\imath = \sqrt{-1}$. It completely determines behavior and properties of the probability distribution of the random variable X: As a matter of fact, it is possible to show that the characteristic function univocally "characterizes" the cumulative distribution function and vice versa (note that φ can be seen as the Fourier transform of the p.d.f. with the sign of $\imath$ changed, which can therefore be deduced from it by inverse Fourier transform). In addiction, if the moments of the random variable exist, the characteristic function computed in $-\imath\, u$ correspond to the moment-generating function.

A.2 Stochastic processes and their characterization

As functions are the natural extension of number, stochastic processes are the one for random variables. In the following, we will mainly refer to [12].

Definition A.2.1 *Given a probability space* $(\Omega, \mathcal{F}, \mathbf{P})$ *and a set* T*, which is a close subset of* $[0, +\infty)$*, we define a continuous-time real valued* stochastic process *as a family of random variables* $\{X(t)\}_{t \in T}$ *defined on* $(\Omega, \mathcal{F}, \mathbf{P})$ *such that*[3]

$$X(t) : \Omega \to \mathbb{R}^n \qquad \forall t \in T\,. \tag{A.6}$$

We will often use the shorter notation $\{X(t)\}$ instead of the previous notation. Examples of T are a closed interval $[a, b] \subseteq [0, +\infty)$ or the real number line itself $\mathbb{R}_0^+$. If we assume $T \subseteq \mathbb{N}$, the process becomes a discrete-time stochastic process[4].

Given a process $\{X(t)\}$, the map $t \mapsto X(t, \omega)$ defined over T with values in $\mathbb{R}^n$, is called *path* of $\omega \in \Omega$. Thus the paths of a process $\{X(t)\}$ are all the functions $t \mapsto X(t, \omega)$ obtained on varying $\omega \in \Omega$. Moreover, a process $\{X(t)\}$ is said to be *continuous* if its paths are almost surely continuous, that is, there exists a event $A' \in \mathcal{F}$ with $\mathbf{P}(A') = 0$, such that, if $\omega \notin A'$, the function $t \mapsto X(t, \omega)$ is continuous on T (but, pay attention, we are

[3] In most cases, in the following dissertation we will deal with processes such that $n = 1$, which means $X(t) : \Omega \to \mathbb{R}$. Thus, if not specifically said, we are relating to them.

[4] Discrete stochastic process, which we basically will not deal with, are usually denoted by $\{X_n\}_{n \in \mathbb{N}}$ in order to distinguish them from the continuous ones (how it is usually done in characterizing functions $f(t)$ as opposed to sequences a_n).

not dealing with the continuity of the parameter t, which is assumed to be continuous by the definition).

Consider a stochastic process $\{X(t)\}$, we fix a instant $t_1 \in T$ in which we set $X(t_1) = X_1$. This is a random variable whose c.d.f is defined as:

$$F_{X(t_1)}(x_1) = \Pr\{X_1 \leq x_1\}\,.$$

This is said *fist order distribution* of $\{X(t)\}$. Similarly, given $t_1, t_2 \in T$ and $X(t_1) = X_1$, $X(t_2) = X_2$ they represent two random variables, whose joint distribution is

$$F_{X(t_1),X(t_2)}(x_1, x_2) = \Pr\{X_1 \leq x_1, X_2 \leq x_2\}$$

which is named *second order distribution*. More generally, we define *n–th order distribution* the following

$$F_{X(t_1),X(t_2),...,X(t_n)}(x_1, x_2, ..., x_n) = \Pr\{X_1 \leq x_1, X_2 \leq x_2, ..., , X_n \leq x_n\}\,.$$

At first glance, the complete characterization of $\{X(t)\}$ should require the knowledge of all the distributions as $n \to \infty$. However, in few cases the amount of information required will be less demanding.

We now introduce a series of basic functions of a stochastic process which sometimes can be sufficient to wholly describe it. First we introduce the one-variable function *mean* of the process

$$\mu_X(t) := \mathbb{E}[X(t)] \tag{A.7}$$

where $X(t)$ is considered as a random variable for a given t. In general, $\mu_X(t)$ is a function of the time. A measure of (linear) dependence between the variables of $\{X(t)\}$ is the two-variable *autocorrelation function*, which is defined by:

$$\mathcal{R}_X(t, s) := \mathbb{E}[X(t) \cdot X(s)]\,. \tag{A.8}$$

This function results symmetric ($\mathcal{R}_X(t, s) = \mathcal{R}_X(s, t)$) and thus knowing it only for $t \geq s$ is enough to describe all. Another two-variable function related to the previous ones is the *autocovariance function* given by:

$$\begin{aligned} \mathcal{C}_X(t, s) :&= \mathbb{E}\big\{[X(t) - \mu_X(t)] \cdot [X(s) - \mu_X(s)]\big\} = \\ &= \mathcal{R}_X(t, s) - \mu_X(t) \cdot \mu_X(s)\,. \end{aligned} \tag{A.9}$$

Even this function is symmetric and, when both $\mu_X(t) = \mu_X(s) = 0$, the relation $\mathcal{C}_X(t, s) = \mathcal{R}_X(t, s)$ becomes obvious. In particular we set

$$\mathcal{C}_X(t, t) = \mathbb{E}\big\{[X(t) - \mu_X(t)]^2\big\} = \mathbb{V}[X(t)] = \sigma_X^2(t) \tag{A.10}$$

which is the variance of the process $\{X(t)\}$. The last quantity we introduce is the two-variable *autocorrelation coefficient function* which is defined as:

$$\rho_X(t,s) := \frac{\mathcal{C}_X(t,s)}{\sigma_X(t)\cdot\sigma_X(s)}\,. \tag{A.11}$$

See that, from the definition, we have $\rho_X(t,t) = 1$ and $|\rho_X(t,s)| \leq 1$.

These functions are usually used to describe some process features and track feasible "repeats" of it, in the sense of likenesses between the process and a time translation of itself. For instance, if $\{X(t)\}$ is roughly periodic with period $k \in \mathbb{R}$, $\rho_X(t+k,t)$ will be higher in modulus than the other values of $\rho_X(t,s)$ (with the exception of $\rho_X(t,t)$ which is always equal to one).

If a process $\{X(t)\}$ possesses a particular probabilistic structure, a less bits of information are required in order to characterize it entirely. Indeed some simple stochastic processes are totally described by the only first and second order distributions. A stochastic process $\{X(t)\}_{t\in T}$ is said to be *strictly stationary* if and only if

$$F_{X(t_1),X(t_2),...,X(t_n)}(x_1,x_2,...,x_n) = F_{X(t_1+\tau),X(t_2+\tau),...,X(t_n+\tau)}(x_1,x_2,...,x_n)$$

for all $t_1,t_1,...,t_n \in T$ and $\tau \in \mathbb{R}$. It follows that the distribution of the process is not influenced by the translation of the time-origin, thus $\{X(t)\}$ and $\{X(t+\tau)\}$ have the same distribution; this circumstance can also be indicated with the notation $X(t) \overset{d}{\sim} X(t+\tau)$, considering them as two *identically distributed* random variables calculated in t and $t+\tau$ respectively. As a matter of fact, for the first order distribution, we have

$$F_{X(t)}(x) = F_{X(t+\tau)}(x) = F_X(x)$$

and thus

$$\mathbb{E}\Big[X(t)^k\Big] = m_k$$

so neither the c.d.f of the process nor its moments of any order, if they exist, will depend on t.

If the condition for the strictly stationarity is not valid for all the values of n, but only for $m \leq n$, we have just m–th order distributions equals to each other and the process is m–th order stationary. Obviously, m–th order stationarity implies lower order stationarity. If $m = 2$, the process $\{X(t)\}_{t\in T}$ is said *weakly stationary*. For such a process, the mean and the autocorrelation function do not depend on t. This implies that the first one is a constant and the second one depend only on one variable instead of two. So $\mu_X(t) = \mu_X$ and $\mathcal{R}_X(t+\tau,t) = \mathcal{R}_X(\tau)$. It follows that even all the other functions defined above become one-variable functions.

In addiction, a stochastic process $\{X(t)\}$ is *degenerate* if, for all $t \in T$, $X(t) \equiv 0$ almost surely.

For a random process, if all the random variables $X(t_i)$, with $t_i \in T$ $\forall i = 1, 2, ..., n$, are independent random variables, that is

$$F_{X(t_1),X(t_2),...,X(t_n)}(x_1, x_2, ..., x_n) = \prod_{i=1}^{n} F_{X(t_i)}(x_1)$$

thus we say that $\{X(t)\}_{t\in T}$ is a *independent stochastic process*. This mean that the first order distribution is sufficient to characterize the whole process. In addiction we say that a process has *independent increments* if, for all $0 = t_0 < t_1 < t_2 < ... < t_n$

$$X(t_1) - X(t_0)\,,\; X(t_2) - X(t_1)\,,\; ...\,,\; X(t_n) - X(t_{n-1})$$

are independent. On the other hand, if the stochastic process has independent increments and if $X(t)-X(s) \overset{d}{\sim} X(t+\tau)-X(s+\tau)$, for all $s < t, \tau \in T$, it exhibits *independent and stationary increments*. It can be trivially proved that, granted $X(0) \equiv 0$, such a process gets the following properties:

$$\mu_X(t) = \mu_X(1) \cdot t$$

$$\sigma_X^2(t) = \sigma_X^2(1) \cdot t$$

$$\mathbb{V}[X(t) - X(s)] = \sigma_X^2(1) \cdot (t - s)$$

$$\mathcal{C}_X(t, s) = \sigma_X^2(1) \cdot \min\{t, s\}\,.$$

Looking at the first properties, we see that processes with independent and stationary increments are not stationary. Examples of such processes are Poisson Process (that will not be treated in this text) and Wiener Process.

Eventually, we define another central class of process. A real stochastic process $\{X(t)\}_{t\in T}$ is said *Gaussian* if, for all $t_1, t_2, ..., t_n \in T$, all the random variables which constitute the random vector $(X(t_1), X(t_2), ..., X(t_n))$ are normal, that is, any finite linear combination of the $X(t_i)$, with $i = 1, 2, ..., n$, is a random variable normally distributed. A basic property of such processes is that their finite-dimensional properties are univocally determined by the mean $\mu_X(t)$ and the autocovariance function $\mathcal{C}_X(t, s)$. Such process is usually indicated as

$$X(t) \overset{d}{\sim} \mathcal{N}[\mu_X(t), \mathcal{C}_X(t, s)]$$

highlighting the connection with the distribution of the random variable X computed in the instant t. Moreover, it is possible to prove that, if the normal variables $X(t_1), X(t_2), ..., X(t_n)$ are mutually uncorrelated, they are also mutually independent and vice versa.

A.3 Self-affine processes

In the following section, we will concentrate on a property which usually is not treated in most texts about stochastic process theory. A more general geometric discussion of self-affinity and self-similarity will be given in Chapter 2.

Definition A.3.1 *Granted $T = \mathbb{R}_0^+$, a real stochastic process is called* self-affine *with a positive scaling index H (shortly H–*s.a.*), if given $a > 0$ the following condition holds:*

$$\{X(a \cdot t)\}_{t \geq 0} \overset{d}{\sim} \left\{a^H \cdot X(t)\right\}_{t \geq 0} .$$

The H index is said Hurst exponent of the process, whose name is due to its former user. The first consequence to be taken into account is that a non degenerate stochastic process cannot be both stationary and self-affine. As a matter of fact, if we consider a stationary stochastic process, it follows:

$$\{X(t)\}_{t \geq 0} \overset{d}{\sim} \{X(a \cdot t)\}_{t \geq 0} .$$

If, in addiction, it had to be self-affine as well, the following condition should be satisfied

$$\{X(a \cdot t)\}_{t \geq 0} \overset{d}{\sim} \left\{a^H \cdot X(t)\right\}_{t \geq 0}$$

but taking the limit for $a \to +\infty$, the process diverges (it becomes infinitely large) so not being stationary. Although, given a stationary process it is always possible to obtain a self-affine process and vice versa thought simple transformations. Let we consider a H–s.a. process $\{X(t)\}_{t \geq 0}$, so the process

$$\{Y(t)\}_{t \geq 0} = e^{tH} \cdot \left\{X\left(e^t\right)\right\}_{t \geq 0}$$

results stationary. On the other hand, if $\{Y(t)\}_{t \geq 0}$ is a stationary process, the process

$$\{X(t)\}_{t \geq 0} = t^H \cdot \{Y\left(\ln(t)\right)\}_{t \geq 0}$$

is a H–s.a. process.

Now we will present a more general class of processes being self-affine and having stationary increments, even indicated as H–s.a.s.i. process.

Proposition A.3.2 *If a finite-variance stochastic process is* H–s.a.s.i.*, if it owns the following properties:*

1. $X(0) = 0$ *almost surely.*

   Proof: As a matter of fact, for all $a > 0$, it ensues

   $$\{X(a \cdot 0)\} \overset{d}{\sim} \left\{0^H \cdot X(0)\right\} .$$

□

2. *If* $H \neq 1 \Rightarrow \mathbb{E}[X(t)] = 0$.

   Proof: Since $\mathbb{E}[X(2t)] = 2^H \cdot \mathbb{E}[X(t)]$ holds, from the stationarity of the increments and from property 1, we have:

$$\begin{aligned} 2^H \cdot \mathbb{E}[X(t)] = \mathbb{E}[X(2t)] &= \mathbb{E}[X(2t)] - \mathbb{E}[X(t)] + \mathbb{E}[X(t)] = \\ &= \mathbb{E}[X(2t) - X(t)] + \mathbb{E}[X(t) - X(0)] = \\ &= 2 \cdot \mathbb{E}[X(t)] \,. \end{aligned}$$

   Thus $\mathbb{E}(X_t) = 0$ is necessary.

□

3. *It is possible to extend the process time to the whole real line* ($T \subseteq \mathbb{R}$), *since* $\{X(-t)\}_{t \geq 0} \overset{d}{\sim} \{-X(t)\}_{t \geq 0}$.

   Proof: Considering property 1 and the stationarity of the increments, we have:

$$\{X(-t)\}_{t \geq 0} = \{X(-t) - X(0)\}_{t \geq 0} \overset{d}{\sim} \{X(0) - X(t)\}_{t \geq 0} = \{-X(t)\}_{t \geq 0} \,.$$

   Hence $\{X(t)\}_{t \in \mathbb{R}}$.

□

4. *Let* $\sigma_X^2(1) = \sigma^2$, *then the variance of the process is equal to*

$$\sigma_X^2(t) = \sigma^2 \cdot |t|^{2H} \,.$$

   Proof: It follows from properties 2, 3 and self-affine definition.

$$\begin{aligned} \sigma_X^2(t) = \mathbb{E}\left[X(t)^2\right] &= \mathbb{E}[X(t) \cdot X(t)] = \\ &= \mathbb{E}\left[X(|t| \cdot \operatorname{sgn}(t)) \cdot X(|t| \cdot \operatorname{sgn}(t))\right] = \\ &= \mathbb{E}\left[X(\operatorname{sgn}(t))^2\right] \cdot |t|^H \cdot |t|^H = \mathbb{E}\left[X(\operatorname{sgn}(t))^2\right] \cdot |t|^{2H} \,. \end{aligned}$$

   This quantity is equal to (due to property 3 and to the positivity of H):

$$\begin{aligned} \mathbb{E}\left[X(\operatorname{sgn}(t))^2\right] \cdot |t|^{2H} &= \begin{cases} \mathbb{E}\left[X(1)^2\right] \cdot t^{2H} & \text{if } t > 0 \\ 0 & \text{if } t = 0 \\ \mathbb{E}\left[X(-1)^2\right] \cdot (-t)^{2H} & \text{if } t < 0 \end{cases} \\ &= \begin{cases} \mathbb{E}\left[X(1)^2\right] \cdot t^{2H} & \text{if } t > 0 \\ 0 & \text{if } t = 0 \\ \mathbb{E}\left[X(1)^2\right] \cdot (-t)^{2H} & \text{if } t < 0 \end{cases} \\ &= \sigma^2 \cdot |t|^{2H} \,. \end{aligned}$$

□

   Moreover, the process is said *standard*, if we have $\sigma^2 = 1$.

Note that, if we only consider positive times $t \geq 0$, the previous expression becomes:

$$\sigma_X^2(t) = \sigma^2 \cdot t^{2H} .$$

5. *The autocovariance function is equal to*

$$\mathcal{C}_X(t,s) = \frac{\sigma^2}{2} \cdot \left(|t|^{2H} + |s|^{2H} - |t-s|^{2H} \right) .$$

Proof: It follows from properties 2 and 4 and from the stationarity of the increments. Indeed, given $H \neq 1$, since the mean $\mathbb{E}[X(t)]$ is null, we have:

$$\begin{aligned}
\mathcal{C}_X(t,s) = \mathcal{R}_X(t,s) &= \mathbb{E}[X(t) \cdot X(s)] = \\
&= \frac{2 \cdot \mathbb{E}[X(t) \cdot X(s)]}{2} + \mathbb{E}[X(t)^2] - \mathbb{E}[X(t)^2] + \mathbb{E}[X(s)^2] - \mathbb{E}[X(s)^2] = \\
&= \frac{2 \cdot \left\{ \mathbb{E}[X(t)^2] - \mathbb{E}[X(t)^2] \right\} + 2 \cdot \left\{ \mathbb{E}[X(s)^2] - \mathbb{E}[X(s)^2] \right\} + 2 \cdot \mathbb{E}[X(t) \cdot X(s)]}{2} = \\
&= \frac{\mathbb{E}[X(t)^2] + \mathbb{E}[X(s)^2] - \left\{ \mathbb{E}[X(t)^2] + \mathbb{E}[X(s)^2] - 2 \cdot \mathbb{E}[X(t) \cdot X(s)] \right\}}{2} = \\
&= \frac{\mathbb{E}[X(t)^2] + \mathbb{E}[X(s)^2] - \mathbb{E}\left[X(t)^2 + X(s)^2 - 2 \cdot X(t) \cdot X(s) \right]}{2} = \\
&= \frac{\mathbb{E}[X(t)^2] + \mathbb{E}[X(s)^2] - \mathbb{E}\left\{ [X(t) - X(s)]^2 \right\}}{2} = \\
&= \frac{|t|^{2H} \cdot \sigma^2 + |s|^{2H} \cdot \sigma^2 - |t-s|^{2H} \cdot \sigma^2}{2} = \\
&= \frac{\sigma^2}{2} \cdot \left(|t|^{2H} + |s|^{2H} - |t-s|^{2H} \right) .
\end{aligned}$$

□

Furthermore, if we we take into account only positive times $s, t \geq 0$, the covariance function is:

$$\mathcal{C}_X(t,s) = \frac{\sigma^2}{2} \cdot \left(t^{2H} + s^{2H} - |t-s|^{2H} \right) .$$

6. *If $\{X_t\}_{t \geq 0}$ is a H–s.a.s.i. process, it must be $0 < H < 1$.*

Proof: The bound of the Hurst exponent arises form the request of increments' stationarity. Indeed:

$$\begin{aligned}
&\mathbb{E}\big[|X(2) + X(1) - X(1)|\big] \leq \mathbb{E}\big[|X(2) - X(1)|\big] + \mathbb{E}\big[|X(1)|\big] \\
&\mathbb{E}\big[|X(2)|\big] \leq \mathbb{E}\big[|X(1) - X(0)|\big] + \mathbb{E}\big[|X(1)|\big] \\
&\mathbb{E}\big[|X(2 \cdot 1)|\big] \leq \mathbb{E}\big[|X(1)|\big] + \mathbb{E}\big[|X(1)|\big] \\
&2^H \cdot \mathbb{E}\big[|X(1)|\big] \leq 2 \cdot \mathbb{E}\big[|X(1)|\big] \\
&2^H \leq 2 .
\end{aligned}$$

Thus it ensues:

$$H \leq 1 .$$

Since H is by definition a positive number, the final condition for increments' stationarity is[5]:

$$0 < H < 1\,.$$

□

A.4 Filtrations, martingales and semimartingales

An increasing collection $\mathcal{F}_t = \{\mathcal{F}_t\}_{t\in T}$ of sub σ-algebras of $\mathcal{F}$ containing the null measure events is said *filtration* in $(\Omega, \mathcal{F}, \mathbf{P})$. That is a family of σ-algebras $\mathcal{F}_s \subseteq \mathcal{F}_t \subset \mathcal{F}$, if $s \leq t < +\infty$ with $t, s \in T$. Intuitively $\mathcal{F}_t$ gathers all the known information (or events) until the time t. Below, we always will consider *complete filtration* which can be defined as follow. Granted a set $\Phi = \{F \in \mathcal{F} : \mathbf{P}(F) = 0\}$, we will assume that $\Phi \in \mathcal{F}_t$, for all $t \in T$ (so the null measure event is a element of the filtration itself). In addiction, it is always possible to extend a given filtration to get a complete one.

A probability space in which a complete filtration is also defined, that is a sort of object $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\in T}, \mathbf{P})$, is named *filtered probability space*. Assigned a process $\{X(t)\}$ on a filtered probability space, it is said to be *adapted to the filtration* or $\mathcal{F}_t-$ *adapted*, if for all $t \in T$, $X(t) : (\mathcal{F}_t, \Omega) \to \mathbb{R}^n$ is measurable. A process is always adapted to its *natural filtration* $\left\{\mathcal{F}_t^X\right\}_{t\in T}$, where $\mathcal{F}_t^X = \sigma(X(s), s \leq t)$ is the smallest σ-algebra which makes all the random variable $X(s)$ measurable. Note that $\{X(t)\}$ results to be $\mathcal{F}_t-$ adapted if and only if:

$$\mathcal{F}_t^X \subseteq \mathcal{F}_t, \quad t \in T\,.$$

Thus $\{X(t)\}$ it is said to be $\mathcal{F}_t-$ measurable for all $t \in T$.

Given the probability space $(\Omega, \mathcal{F}, \mathbf{P})$, let $\mathcal{G}$ be a sub-σ-algebra of $\mathcal{F}$. It is possible to prove that for all Lebesgue integrable random variable X defined on Ω, there is a real random variable Z which is $\mathcal{G}-$ measurable (that is, $Z : (\Omega, \mathcal{G}) \to \mathbb{R}$ measurable), such that, $\forall G \in \mathcal{G}$, the following relation is valid:

$$\int_G X(\omega)\,\mathbf{P}(\mathrm{d}\omega) = \int_G Z(\omega)\,\mathbf{P}(\mathrm{d}\omega)\,. \tag{A.12}$$

[5]Note that, if H were null, the definition of the process would become inconsistent since:

$$\{X(a\cdot t)\}_{t\geq 0} \overset{d}{\sim} \left\{a^0 \cdot X(t)\right\}_{t\geq 0}$$

$$\{X(a\cdot t)\}_{t\geq 0} \overset{d}{\sim} \{X(t)\}_{t\geq 0}\,.$$

That way, becoming a stationary process (and so non self-affine). Moreover, if H were equal to 1, we would obtain an (isotropically) *self-similar* stochastic process, but the property 2 would not hold anymore, letting the process having a non-zero mean (the process becomes degenerate, being a straight line with random slope).

This is the same as requiring that $\mathbb{E}(X \cdot Y) = \mathbb{E}(Z \cdot Y)$ for all random variable Y which is $\mathcal{G}-$ measurable and limited. The variable Z is not unique; nevertheless, if Z_1 and Z_2 are two $\mathcal{G}-$ measurable variables such that the previous relation holds, we have $Z_1 = Z_2$ almost surely. Moreover we call Z *conditional expectation* of X given $\mathcal{G}$, indicating it with $\mathbb{E}(X|\,\mathcal{G})$. Intuitively, the conditional expectation $\mathbb{E}(X|\,\mathcal{G})$ is the $\mathcal{G}-$ measurable random variable which best approximates X[6].

Definition A.4.1 *A real $\mathcal{F}_t$ – adapted stochastic process $\{X(t)\}_{t\in T}$ is said to be a* submartingale (supermartingale), *if it is integrable for all $t \in T$ ($\mathbb{E}\left[|X(t)|\right] < \infty$) and the following properties, respectively, are valid:*

$$\mathbb{E}[X(t)|\,\mathcal{F}_s] \geq X(s) \quad \text{a.s.} \qquad \Big(\mathbb{E}[X(t)|\,\mathcal{F}_s] \leq X(s) \quad \text{a.s.}\Big)$$

for all $t, s \in T$ with $s \leq t$. Thus, a real $\mathcal{F}_t$ – adapted stochastic process is said to be a martingale *if it is both a submartingale and a supermartingale, that is:*

$$\mathbb{E}[X(t)|\,\mathcal{F}_s] = X(s) \quad \text{a.s.} \tag{A.13}$$

for all $t, s \in T$ with $s \leq t$.

This condition can also be expressed as[7]:

$$\mathbb{E}[X(t) - X(s)|\,\mathcal{F}_s] = 0 \quad \text{a.s.}$$

Given $t \geq 0$, from the definition we have that the expected value of such a process does not depend on the time. To show it easily, we can take into account the filtration $\mathcal{F}_0$, giving:

$$\mathbb{E}[X(t)] = \mathbb{E}\big\{\mathbb{E}[X(t)|\,\mathcal{F}_0]\big\} = \mathbb{E}[X(0)]\,.$$

Thus a martingale has always a *constant* mean through time.

Martingales can be defined even when the probability space is not provided with a filtration: In this case we suppose the martingale based on

[6] In addiction, it satisfies the following properties:

- if X is $\mathcal{G}$–measurable and $X \cdot Y$ is integrable $\Rightarrow \mathbb{E}(X \cdot Y|\,\mathcal{G}) = X \cdot \mathbb{E}(Y|\,\mathcal{G})$ a.s. (measurability);
- if X is independent given $\mathcal{G} \Rightarrow \mathbb{E}(X|\,\mathcal{G}) = \mathbb{E}(X)$ a.s. (independence);
- if $\mathcal{H} \subseteq \mathcal{G} \Rightarrow \mathbb{E}[\mathbb{E}(X|\,\mathcal{G})|\,\mathcal{H}] = \mathbb{E}(X|\,\mathcal{H})$ a.s. (refinement).

[7] This definition of martingale is related to the one of *fair game*. It is indeed defined as a sequence (hence discrete) of bets $\{G_n\}_{n\in\mathbb{N}}$ whose single expected value are null, that is:

$$\mathbb{E}(G_n|\,\mathcal{F}_{n-1}) = 0 \quad \text{with} \quad n = \{1, 2, ...\}\,.$$

Thus a (discrete) martingale can be naively characterized as a process whose increments are the "gains" of a fair game.

the natural filtration $\mathcal{F}_t^X := \sigma\left(\{X(u)\}_{u\in[0,t]\cap T}\right)$. If we need to highlight a particular filtration, we will say that the process is a $\{\mathcal{F}_t\}_{t\in T}$ – martingale.

Finally, we introduce a more general concept usually involved in stochastic analysis. We write $\mathcal{M}$ for the set of all càdlàg[8] martingales. First of all we need to define the *stopping time*.

Definition A.4.2 *Let $\{X(t)\}_{t\in[0,+\infty)}$ be a stochastic process. A* stopping time *with respect to $\{X(t)\}$ is a random time (a discrete random variable) such that for each $t\in[0,+\infty)$ the event $\{\tau=t\}$ is completely determined by (at most) the total information known up to time t, that is $\{X(s)\}_{s\in[0,t]}$*[9].

The following result is fundamental. For the proof, we refer to [12].

Theorem A.4.3 (**Optimal stopping theorem**) *Let $\{X(t)\}_{t\in[0,+\infty)}$ be a càdlàg $\mathcal{F}_t$ – adapted integrable process. Then the following are equivalent:*

(a) *$\{X(t)\}$ is a martingale;*

(b) *$\{X(\min\{\tau,t\})\}$ is a martingale for all bounded stopping times τ;*

(c) *For all bounded stopping times τ and υ, $\mathbb{E}[X(\tau)|\,\mathcal{F}_\upsilon]=X(\min\{\tau,\upsilon\})$;*

(d) *$\mathbb{E}[X(\tau)]=\mathbb{E}[X(0)]$ for all bounded stopping times τ.*

It is also the case that $\mathcal{M}$ is *stable under stopping*. This observation leads us to define a slightly more general class of processes, called *local martingales*[10].

Definition A.4.4 *A càdlàg $\mathcal{F}_t$ – adapted process $\{M(t)\}_{t\in[0,+\infty)}$ is a* local martingale*, if there exist a sequence of stopping times $\{\tau_n\}_{n\in\mathbb{N}}$ with $\lim_{n\to\infty}\tau_n=\infty$ a.s., such that the stopped process $\{M(\min\{t,\tau_n\})\}_{t\in[0,+\infty)}$ is a martingale for all $n\in\mathbb{N}$.*

Such a process is often indicated as $\{M(t)\}_{t\in[0,+\infty)}\in\mathcal{M}_{loc}$. In particular, by Theorem A.4.3, it follows $\mathcal{M}_{loc}\subseteq\mathcal{M}$. Eventually we have all the tools to define a *semimartingale*.

---

[8] A function $f:\mathbb{R}\supset A\to B$ is said to be a càdlàg function (from the French *continue à droite, limite à gauche*) if, for every $x\in A$

$$\exists\lim_{x\to x_0^+}f(x)=f(x_0)\quad\text{and}\quad\exists\lim_{x\to x_0^-}f(x)\,.$$

[9] If $X(t)$ denotes the price of a stock at time t and τ denotes the time at which we will sell the stock (or buy the stock), then our decision to sell (or buy) the stock at a given time can only depend on the information known at that time (not on future information). The time at which one might exercise an option is yet again another example.

[10] A martingale can be interpreted as the fortune of a player in a fair game. A local martingale which is not always a martingale, on the other hand, is the fortune of a player in a game which looks locally fair: Unfortunately, this is only because there are going to be times of huge increases of X followed by an eventual ruin. Overall, the expected fortune decreases. A local martingale is thus something akin to a bubble in the market.

Definition A.4.5 *A continuous* semimartingale $\{X(t)\}_{t\in[0,+\infty)}$ *is an adapted continuous process which may be written as*

$$X(t) = X(0) + M(t) + A(t) \tag{A.14}$$

with $M(0) = A(0) = 0$ *almost surely, where* $\{M(t)\}_{t\in[0,+\infty)} \in \mathcal{M}_{loc}$ *is a local martingale and* $\{A(t)\}_{t\in[0,+\infty)}$ *is a càdlàg adapted process of locally bounded variation*[11] .

The class of semimartingales remains invariant under several transformations such as stopping, localization, change of time, change of filtration, absolutely continuous change of measure, etc. More importantly, semimartingales represent the largest class of processes with respect to which the Itō integral can be defined (see Section 1.2). Submartingales, supermartingales, and hence martingales, together represent a subset of the semimartingales.

[11] We say that a process $\{A(t)\}$ has *locally* bounded variation, if in $[a,b] \subset T$, given a partition $\Pi = \{t_0, t_1, ..., t_n\}$ with $a = t_0$ and $b = t_n$, we have

$$\sup_{\Pi} \left\{ \sum_{i=1}^{n} |A(t_i) - A(t_{i-1})| \right\} < \infty \qquad \text{a.s.}$$

Moreover, $\{A(t)\}$ is said to be of bounded variation if the above inequality is valid for all bounded intervals $[a,b] \subset T$.

Appendix B

Appendix

B.1 Elements of metric topology

The second appendix is devoted to the exposition of few concepts involving metric topology. They are very useful for the comprehension of the subjects of Chapter two. Most of the definition and theorems are taken from [35] and [36].

Definition B.1.1 *A set X is said to be a* vector space *over a field*[1] Λ *if the following two operation are defined for all its elements:*

$\mathcal{V}_1$. *There exists an application*[2] $X \times X \to X$, *denoted with the symbol "+" which is named* addition, *such that:*

1. $\forall \mathbf{x}, \mathbf{y} \in X : \mathbf{x} + \mathbf{y} = \mathbf{y} + \mathbf{x}$ (commutativity of addition);

[1] Given a set A, it is said to be a *field* if, $\forall a, b, c \in A$, there exist two operations, sum (indicated by "+") and multiplication (indicated by "$\cdot$"), for which the following properties hold:

1. $a + b \in A$ and $a \cdot b \in A$ (closure of A under addition and multiplication);
2. $(a + b) + c = a + (b + c)$ and $(a \cdot b) \cdot c = a \cdot (b \cdot c)$ (associativity of addition and multiplication);
3. $a + b = b + a$ and $a \cdot b = b \cdot a$ (commutativity of addition and multiplication);
4. $\exists 0 \in A : a + 0 = a$ and $\exists 1 \in A : a \cdot 1 = a$ (existence of additive and multiplicative identity elements);
5. $\exists -a \in A : a + (-a) = 0$ and $\exists a^{-1} \in A : a \cdot a^{-1} = 1$ (existence of additive inverses and multiplicative inverses);
6. $a \cdot (b + c) = a \cdot b + a \cdot c$ (distributivity of multiplication over addition).

[2] The symbol "$\times$" is the so-called *Cartesian product*. Given n sets $X_1, X_2, ..., X_n$ the n-ary Cartesian product is defined as

$$X_1 \times X_2 \times ... \times X_n := \{(x_1, x_2, \ldots, x_n) : x_i \in X_i\} .$$

2. $\forall \mathbf{x}, \mathbf{y}, \mathbf{z} \in X : (\mathbf{x}+\mathbf{y})+\mathbf{z} = \mathbf{x}+(\mathbf{y}+\mathbf{z})$ (associativity of addition);
3. $\exists \mathbf{0} \in X : \forall \mathbf{x} \in X \Rightarrow \mathbf{x}+\mathbf{0} = \mathbf{x}$ (existence of the identity element of addition);
4. $\exists -\mathbf{x} \in X : \forall \mathbf{x} \in X \Rightarrow \mathbf{x}+(-\mathbf{x}) = \mathbf{0}$ (existence of the inverse element of addition).

$\mathcal{V}_2$. *There exists an application* $\Lambda \times X \to X$, *denoted with* $\lambda\mathbf{x}$, *if* $\lambda \in \Lambda$ *and* $\mathbf{x} \in X$, *which is named* multiplication for the elements of Λ, *such that* $\forall \lambda, \mu \in \Lambda$, $\forall \mathbf{x}, \mathbf{y} \in X$:

1. $\lambda(\mu\mathbf{x}) = \mu(\lambda\mathbf{x})$ (compatibility of scalar multiplication with field multiplication);
2. $1\mathbf{x} = \mathbf{x}$ (identity element of scalar multiplication);
3. $(\lambda+\mu)\mathbf{x} = \lambda\mathbf{x}+\mu\mathbf{x}$ (distributivity of scalar multiplication with respect to field addition);
4. $\lambda(\mathbf{x}+\mathbf{y}) = \lambda\mathbf{x}+\lambda\mathbf{y}$ (distributivity of scalar multiplication with respect to vector addition).

The elements of X are named *vectors* and are usually denoted with lower case bold-type Latin letters, whereas the elements of the field Λ are called *scalars* and are denoted with lower case Latin/Greek letters. Form properties $\mathcal{V}_1$ and $\mathcal{V}_2$ it easily follows that $\mathbf{0} = 0\mathbf{x}$ and $-\mathbf{x} = (-1)\mathbf{x}$. Moreover the zero of the space and the inverse of all its element are unique.

Definition B.1.2 *A vector* $\mathbf{x}$ *is said to be a* linear combination *of the vectors* $\mathbf{x_1}, \mathbf{x_2}, ..., \mathbf{x_k}$ *if it can be written as*

$$\mathbf{x} := \lambda_1\mathbf{x_1} + \lambda_2\mathbf{x_2} + ... + \lambda_k\mathbf{x_k}$$

with the scalars $\lambda_1, \lambda_2, ..., \lambda_k$ *named* coefficients *of the combination.*

Definition B.1.3 *We say that* k *vectors* $\mathbf{x_1}, \mathbf{x_2}, ..., \mathbf{x_k}$ *are* linearly dependent *if there are* k *scalars* $\alpha_1, \alpha_2, ..., \alpha_k$ *not all zero such that*

$$\alpha_1\mathbf{x_1} + \alpha_2\mathbf{x_2} + ... + \alpha_k\mathbf{x_k} = \mathbf{0}\,.$$

Otherwise they are said to be linearly independent.

We underline that if the vectors $\mathbf{x_1}, \mathbf{x_2}, ..., \mathbf{x_k}$ are linearly dependent, hence almost one is a linear combination of the other ones. Vice versa, if the vectors $\mathbf{x_1}, \mathbf{x_2}, ..., \mathbf{x_k}$ are linearly independent if and only if $\alpha_1 = \alpha_2 = ... = \alpha_k = 0$.

Definition B.1.4 *A vector space* X *is said a* finite-dimensional space *if there is a finite subset* $B \subseteq X$ *such that:*

1. *The elements of* B *are linearly independent;*

2. *Every vector of X can be obtained as linear combination of the elements of B.*

For these reasons B is named *basis* of the vector space X. It can be proved that *all* the bases have the same cardinality[3]: This number is called *dimension* of the vector space X. That is, if $B = \{\mathbf{b_1}, \mathbf{b_2}, \ldots, \mathbf{b_n}\} \subseteq X$, the latter is a n-dimensional vector space. This fact can also be written as:

$$\dim(X) = n\,.$$

Then we introduce the concept of *metric space*.

Definition B.1.5 *A* metric space *is an ordered pair (X, d) where d is an application $d : X \times X \to \Lambda$ (named* distance *or* metric*) such that, $\forall\, x, y, z \in X$, the following properties hold:*

1. $d(x, y) \geq 0$ (non-negativity)*;*
2. $d(x, y) = 0 \Leftrightarrow x = y$ (identity of indiscernibles)*;*
3. $d(x, y) = d(y, x)$ (simmetry)*;*
4. $d(x, y) \leq d(x, z) + d(y, z)$ (triangle inequality).

Firstly we have to remark that the metric space is the pair (X, d); as a matter of fact, on the same set X might be defined several metrics $d_1, d_2, \ldots$. In that case, the set X with metric d_1 and the same set with metric d_2 are different metric spaces. If X is also a vector space, we deal with *metric vector space*. In that case we have that the elements $x, y, z \in X$ are vectors (and hence written as $\mathbf{x}, \mathbf{y}, \mathbf{z}$). However, note that the structure of vector space and that of metric space can subsist independently from each other.

Definition B.1.6 *A vector space X over the field Λ is said to be* normed *if there exists an application $\|\cdot\| : X \to \Lambda$ such that, $\forall\, \mathbf{x}, \mathbf{y} \in X$ and $\lambda \in \Lambda$, the following properties hold:*

1. $\|\mathbf{x}\| \geq 0$*;*
2. $\|\mathbf{x}\| = 0 \Leftrightarrow \mathbf{x} = \mathbf{0}$*;*
3. $\|\lambda\mathbf{x}\| = |\lambda| \cdot \|\mathbf{x}\|$*;*
4. $\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$*;*
5. $\big|\|\mathbf{x}\| - \|\mathbf{y}\|\big| \leq \|\mathbf{x} - \mathbf{y}\|$.

[3]The cardinality of a countable set A is the number of elements of the set and it is denoted as $\#(A)$. Cantor proved that, given $\#(\mathbb{N}) = \aleph_0$, the *cardinality of the continuum* $\#(\mathbb{R})$ is equal to $2^{\aleph_0} = \aleph_1$.

Note that a normed space is always metric, using such a definition of distance:

$$d(\mathbf{x}, \mathbf{y}) := \|\mathbf{x} - \mathbf{y}\| .$$

However the converse it is not true since, from the properties of the distance listed in Definition B.1.5, it does not follow the third property of Definition B.1.6.

Definition B.1.7 *Let X be vector space over the field Λ. We define* scalar (or inner) product *an application $< \cdot, \cdot > : X \times X \to \Lambda$ such that, $\forall \mathbf{x}, \mathbf{y}, \mathbf{z} \in X$ and $\lambda \in \Lambda$ the following properties hold:*

1. $< \mathbf{x}, \mathbf{x} > \geq 0$;
2. $< \mathbf{x}, \mathbf{x} >= 0 \Leftrightarrow \mathbf{x} = \mathbf{0}$;
3. $< \mathbf{x}, \mathbf{y} >=< \mathbf{y}, \mathbf{x} >$;
4. $< \lambda\mathbf{x}, \mathbf{y} >=< \mathbf{x}, \lambda\mathbf{y} >= \lambda \cdot < \mathbf{x}, \mathbf{y} >$;
5. $< \mathbf{x} + \mathbf{y}, \mathbf{z} >=< \mathbf{x}, \mathbf{z} > + < \mathbf{y}, \mathbf{z} >$.

Another property of the scalar product, which follows form the ones listed above, is the well-known *Cauchy-Schwarz inequality* which states

$$|< \mathbf{x}, \mathbf{y} >| \leq \sqrt{< \mathbf{x}, \mathbf{x} >} \cdot \sqrt{< \mathbf{y}, \mathbf{y} >}$$

and allows us to introduce *a* norm, given by:

$$\|\mathbf{x}\| := \sqrt{< \mathbf{x}, \mathbf{x} >}$$

Thus, given a inner product space, it is "natural" to define the norm in such a way. At the same time, we want to remark that it is *not* the only possible way. As a matter of fact, there exist normed spaces whose norm is not induced by inner product.

On the other hand, the distance defined over the metric space (X, d) can be expressed with the scalar product, that is:

$$d(\mathbf{x}, \mathbf{y}) = \sqrt{< \mathbf{x} - \mathbf{y}, \mathbf{x} - \mathbf{y} >}$$

Hence any inner product space is a normed space (but the converse is not always true). Moreover such a space can become a metric space, defining a distance as above. Sometimes such a space is known as a *pre-Hilbert space*. Note that there are vector spaces that are not inner product spaces (and thus normed spaces neither); at the same time, there exist metric spaces that are not vector spaces. The following example we introduce is the most intuitive vector space, which is both a inner product space and a metric space.

The set $\mathbb{R}^n$, that is given by the Cartesian product

$$\mathbb{R}^n = \underbrace{\mathbb{R} \times \mathbb{R} \times ... \times \mathbb{R}}_{n \text{ times}}$$

whose element are the n-tuples of real numbers, forms an n-dimensional vector space over the field $\mathbb{R}$. As a matter of fact, let $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ and $\mathbf{y} = (y_1, y_2, \ldots, y_n)$ be vectors of $\mathbb{R}^n$, it follows:

$$\lambda \mathbf{x} := (\lambda \cdot x_1, \lambda \cdot x_2, \ldots, \lambda \cdot x_n)$$

$$\mathbf{x} + \mathbf{y} := (x_1 + y_1, x_2 + y_2, \ldots, x_n + y_n)$$

$$\mathbf{0} = (0, 0, \ldots, 0)$$

$$-\mathbf{x} = (-x_1, -x_2, \ldots, -x_n)$$

Given $\mathbf{e_1} = (1, 0, \ldots, 0)$, $\mathbf{e_2} = (0, 1, \ldots, 0)$, $\ldots$, $\mathbf{e_n} = (0, 0, \ldots, 1)$, it is easy to note that the set $E = \{\mathbf{e_1}, \mathbf{e_2}, \ldots, \mathbf{e_n}\} \subseteq \mathbb{R}^n$ is a basis (the canonical one) for $\mathbb{R}^n$ and then

$$\dim(\mathbb{R}^n) = n\,.$$

It is obvious that $\mathbb{R}^n$ is a pre-Hilbert space. As a matter of fact, given $\mathbf{x} = (x_1, x_2, \ldots, x_n), \mathbf{y} = (y_1, y_2, \ldots, y_n) \in \mathbb{R}^n$, the scalar product can be defined as the application $< \cdot, \cdot > : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ such that:

$$< \mathbf{x}, \mathbf{y} > := \sum_{i=1}^{n} x_i \cdot y_i\,.$$

It is easy to prove that such a definition of scalar product satisfies the essential properties given above. Moreover, the norm of a vector $\|\cdot\| : \mathbb{R}^n \to \mathbb{R}$ can be written as:

$$\|\mathbf{x}\| = \sqrt{\sum_{i=1}^{n} x_i^2}\,.$$

Eventually, through such a norm, the distance between two vectors become the application $d : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ defined in the following way

$$d(\mathbf{x}, \mathbf{y}) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}\,.$$

This metric is called *Euclidean distance* and, if we refer to the Euclidean vector space $\mathbb{R}^k$, with $k \leq 3$, it coincides with ordinary notion of distance between two points.

Below, we introduce several necessary notions for the understanding of the following dissertation.

Definition B.1.8 *Given* $\mathbf{x} \in \mathbb{R}^n$ *and* $r \in \mathbb{R}^+$*, we call* n-sphere (or hypersphere) *of radius* r *and center* $\mathbf{x}$ *the set*

$$S(\mathbf{x}, r) := \{\mathbf{y} \in \mathbb{R}^n : d(\mathbf{x}, \mathbf{y}) = r\} .$$

We define open n-ball *of radius* r *and center* $\mathbf{x}$ *the set*

$$B(\mathbf{x}, r) := \{\mathbf{y} \in \mathbb{R}^n : d(\mathbf{x}, \mathbf{y}) < r\} .$$

Moreover we defined closed n-ball *of radius* r *and center* $\mathbf{x}$ *the set*

$$\overline{B}(\mathbf{x}, r) := \{\mathbf{y} \in \mathbb{R}^n : d(\mathbf{x}, \mathbf{y}) \leq r\} .$$

Through these notions, we are able to introduce the concept of *neighborhood.*

Definition B.1.9 *A set* $A \subseteq \mathbb{R}^n$ *is called a* neighborhood *of a point* $\mathbf{x} \in \mathbb{R}^n$ *if there is some open* n*-ball of radius* r *and center* $\mathbf{x}$ *which is contained in* A*.*

Form the definition it follows that even the open n-ball $B(\mathbf{x}, r)$ is a neighborhood of the point $\mathbf{x}$. A more advanced notion is the one of *topology.*

Definition B.1.10 *Let* X *be a set and let* $\mathcal{T}$ *be a family of subsets of* X*. Then* $\mathcal{T}$ *is called a* topology *on* X *if:*

1. *Both the empty set* $\varnothing$ *and* X *itself are elements of* $\mathcal{T}$*;*
2. *Any union of elements of* $\mathcal{T}$ *is an element of* $\mathcal{T}$*;*
3. *Any intersection of finitely many elements of* $\mathcal{T}$ *is an element of* $\mathcal{T}$*.*

If $\mathcal{T}$ *is a topology on* X*, then the pair* $(X, \mathcal{T})$ *is called a* topological space*.*

In the last part of this section, we will examine the topology of $\mathbb{R}^n$ that we have just introduced.

Definition B.1.11 *Given a subset* E *of* $\mathbb{R}^n$*, we define the* (absolute) complement *of* E *the set:*

$$E^c := \mathbb{R}^n \setminus E$$

Furthermore, the following notions are crucial.

Definition B.1.12 *Let* $E \subseteq \mathbb{R}^n$*. A point* $\mathbf{x}$ *is* interior point *of* E *if there is a neighborhood centered at* $\mathbf{x}$ *which is contained in* E*. It is said to be a* exterior point *of* E *if it is a interior point of* E^c*. It is a* boundary point *of* E *if every neighborhood of* $\mathbf{x}$ *contains at least one point in* E *and at least one point in* E^c*.*

The sets of all the interior points and of all boundary points of E *are denoted respectively by* $\overset{\circ}{E}$ *and* ∂E*. We call* closure *of* E *the set*

$$\overline{E} := E \cup \partial E .$$

From this definition, it is obvious that, for every $E \subseteq \mathbb{R}^n$, the following statements are valid:

$$\overset{\circ}{E} \subseteq E \subseteq \overline{E} \qquad \overline{E} = \overset{\circ}{E} \cup \partial E \qquad \partial E = \overline{E} \setminus \overset{\circ}{E}.$$

Definition B.1.13 *Let $E \subseteq \mathbb{R}^n$. A point $\mathbf{x}$ is said to be an* accumulation point *of E if every neighborhoods of $\mathbf{x}$ contains infinitely many points of E. We call the set of all the accumulation points of E* derived set *of E, and we denote it as E'. If a point is not an accumulation point of E, it is called* isolated point.

An accumulation point of E may be or not be in E. It is clear that every interior points of E are also accumulation point of it. If $E' = \varnothing$ (that is there are no accumulation points) E is said to be a *discrete set*. It is obvious that all the points of a discrete set are isolated (but the converse is false). Eventually, if $E' = E$ (that is, the set E is composed by only accumulation points, hence has no isolated points) it is named a *perfect set*.

Definition B.1.14 *A set $E \subseteq \mathbb{R}^n$ is* open *if every $\mathbf{x} \in E$ is an interior point (hence $E = \overset{\circ}{E}$); the set E is* closed *if its complement E^c is open (hence $E = \overline{E}$).*

Of course $\mathbb{R}^n$ is open. However, since also $\varnothing$ is open ad $\mathbb{R}^n = \varnothing^c$, thus $\mathbb{R}^n$ is even closed. The same applies to $\varnothing$. These two sets are the only ones to be both open and closed contemporarily.

Furthermore, the following theorems are valid (for their proof see [35]) and will be very useful in the following.

Theorem B.1.15 *Let $A_1, A_2, \ldots$ be a family, even infinite, of open (closed) sets. Hence the union (intersection) of them is still an open (closed) set. That is*

$$A = \bigcup_{i=1}^{\infty} A_i \qquad \left(A = \bigcap_{i=1}^{\infty} A_i \right)$$

is open (closed).

Theorem B.1.16 *Let $A_1, A_2, \ldots, A_n$ be a finite family of open (closed) sets. Hence the intersection (union) of them is still an open (closed) set. That is*

$$A = \bigcap_{i=1}^{n} A_i \qquad \left(A = \bigcup_{i=1}^{n} A_i \right)$$

is open (closed).

Eventually, we need the following definitions.

Definition B.1.17 *A set $E \subset \mathbb{R}^n$ is a* bounded set *if there exists a neighborhood of the origin which can contain it. A set both bounded and closed is said a* compact.

Definition B.1.18 *A set $E \subset \mathbb{R}^n$ is said to be* connected *if there do not exist open sets F and G such that $F \cup G$ contains E with $E \cap F$ and $E \cap G$ disjoint and non-empty. A set $E \subset \mathbb{R}^n$ is* totally disconnected *if for each distinct $\mathbf{x}, \mathbf{y} \in E$ there exist disjoint open sets F and G such that $\mathbf{x} \in F$, $\mathbf{y} \in G$ and $(E \cap F) \cup (E \cap G) = E$.*

Alternatively speaking, the set E is said to be connected if it is made only of "one piece". On the other hand, the set E is said to be totally disconnected if for any $\mathbf{x}, \mathbf{y} \in E$ and $\mathbf{x} \neq \mathbf{y}$, we have that $\mathbf{x}$ and $\mathbf{y}$ can be separated by the empty set.

Appendix C

Appendix

C.1 The properties of the scaling function $\tau(q)$

In the following we will be interested in analyzing all the relevant properties of the scaling function $\tau(q)$; particularly, we are interested in its zeros and in its concavity/convexity.

Let we take its explicit form

$$\tau(q) = -\log_b \left[\mathbb{E}\left(U^q\right)\right] - 1\,,$$

where U is one of the $\{U_j\}$ i.i.d. random variable with $j = 0, 1, ..., b-1$. Since these variable are discrete[1] the q-th moment can be also written as

$$\tau(q) = -\log_b \left(\sum_{j=0}^{b-1} \Pr\left\{U_j = u_j\right\} \cdot u_j^q \right) - 1$$

where u_j are the values taken by the random variables.

Setting $q = 0$, we get

$$\tau(0) = -\log_b \left(\sum_{j=0}^{b-1} \Pr\left\{U_j = u_j\right\} \cdot u_j^0 \right) - 1 = -\log_b \left(\sum_{j=0}^{b-1} \Pr\left\{U_j = u_j\right\} \right) - 1 =$$
$$= -\log_b (1) - 1 = -1\,,$$

hence the point $(0, -1)$ is the intercept of the scaling function. We are also interested in the first moment of the random variable, which corresponds to the value $q = 1$. Since $\mathbb{E}(U) = \frac{1}{b}$, we find

$$\tau(1) = -\log_b \left(\sum_{j=0}^{b-1} \Pr\left\{U_j = u_j\right\} \cdot u_j \right) - 1 = -\log_b \left(\frac{1}{b}\right) - 1 =$$
$$= -(-1) - 1 = 0\,.$$

[1] The sequence of random variables $\{U_j\}$ is now assumed to be discrete, but with little effort they can be associated with a continuous density. For that eventuality, see [8].

Thus the point $(1,0)$ is one of the zeros of the function $\tau(q)$. On the existence of other zeros, we will discuss later. An alternative proof, based on self-similarity, about the values $\tau(0)$ and $\tau(1)$ has been given in Section 3.3.

We now analyze the derivatives of $\tau(q)$.

$$\begin{aligned}\tau'(q) &= \frac{\mathrm{d}\left[-\log_b\left(\sum_{j=0}^{b-1}\Pr\{U_j=u_j\}\cdot u_j^q\right)-1\right]}{\mathrm{d}q} = \\ &= -\frac{1}{\sum_{j=0}^{b-1}\Pr\{U_j=u_j\}\cdot u_j^q}\cdot\log_b(e)\cdot\sum_{j=0}^{b-1}\Pr\{U_j=u_j\}\cdot u_j^q\cdot\ln(u_j) = \\ &= -\log_b(e)\cdot\frac{\sum_{j=0}^{b-1}\Pr\{U_j=u_j\}\cdot u_j^q\cdot\ln(u_j)}{\sum_{j=0}^{b-1}\Pr\{U_j=u_j\}\cdot u_j^q}\,.\end{aligned}$$

Since $b\in\mathbb{N}\backslash\{0,1\}$, $-\log_b(e)$ is always negative. Moreover, since $U_j\geq 0$ for all j, all the moments of the random variable are positive (they would be exactly equal to zero only in the degenerate case, that is if the random variables were all null), and hence the denominator is positive. The only entity whose sign may vary is the numerator of the fraction, whose positivity/negativity is due to the quantities $\ln(u_j)$. If we consider the microcanical multifractal, that is $0\leq U_j\leq 1$, the logarithms are all negative, making the numerator negative as well. In this case, we find $\tau'(q)\geq 0$, being so *non-decreasing*. However, we have to underline that, if we were working on the canonical measure, which allows $U_j\geq 0$, we would not be able to state, a priori, if the scaling function in increasing or decreasing.

Then, we will study the concavity/convexity of $\tau(q)$:

$$\begin{aligned}\tau''(q) &= \frac{\mathrm{d}^2\left[-\log_b\left(\sum_{j=0}^{b-1}\Pr\{U_j=u_j\}\cdot u_j^q\right)-1\right]}{\mathrm{d}q^2} = \\ &= -\log_b(e)\cdot\frac{\sum_{j=0}^{b-1}\sum_{i=j+1}^{b-1}\Pr\{U_j=u_j\}\cdot\Pr\{U_i=u_i\}\cdot u_j^q\cdot u_i^q\cdot[\ln(u_j)-\ln(u_i)]^2}{\left(\sum_{j=0}^{b-1}\Pr\{U_j=u_j\}\cdot u_j^q\right)^2}\,.\end{aligned}$$

Since the fraction is always positive, the second derivative is always negative (so, despite the monotonicity which may vary depending on the values taken by the random variables U_j, here we have found a more "strong" property). Hence $\tau''(q)<0$ and the scaling function is always *concave*.

The last features we are interested in investigating is the existence of asymptotes for the function $\tau(q)$. In order to know whether they are (and of what type), we have to compute the following limit,

$$\lim_{q\to+\infty}\tau(q)=\lim_{q\to+\infty}-\log_b\left(\sum_{j=0}^{b-1}\Pr\{U_j=u_j\}\cdot u_j^q\right)-1=\begin{cases}+\infty & \text{if } 0\leq U_j\leq 1\\ -\infty & \text{if } U_j\geq 0\end{cases}$$

and

$$\lim_{q\to-\infty}\tau(q)=\lim_{q\to-\infty}-\log_b\left(\sum_{j=0}^{b-1}\Pr\{U_j=u_j\}\cdot u_j^q\right)-1=-\infty\,.$$

Because of these results, we might have at most two *oblique* asymptotes, one for $q\to-\infty$ and one when $q\to+\infty$. Hence we have two compute two limits. The first is

$$\begin{aligned}\lim_{q\to+\infty}\frac{\tau(q)}{q}&=\lim_{q\to+\infty}\frac{-\log_b\left(\sum_{j=0}^{b-1}\Pr\{U_j=u_j\}\cdot u_j^q\right)-1}{q}=\\&=\lim_{q\to+\infty}-\frac{1}{\ln(b)}\cdot\frac{\ln\left(\sum_{j=0}^{b-1}\Pr\{U_j=u_j\}\cdot u_j^q\right)+1}{q}=\\&=\lim_{q\to+\infty}-\frac{1}{\ln(b)}\cdot\frac{\ln\left(\sum_{j=0}^{b-1}\Pr\{U_j=u_j\}\cdot e^{\ln\left(u_j^q\right)}\right)+1}{q}=\\&=\lim_{q\to+\infty}-\frac{1}{\ln(b)}\cdot\frac{\ln\left(\sum_{j=0}^{b-1}\Pr\{U_j=u_j\}\cdot e^{q\cdot\ln(u_j)}\right)+1}{q}\,.\end{aligned}$$

Since, as $q\to+\infty$

$$\ln\left(\sum_{j=0}^{b-1}\Pr\{U_j=u_j\}\cdot e^{q\cdot\ln(u_j)}\right)\sim\ln\left(\Pr\{U_{max}=u_{max}\}\cdot e^{q\cdot\ln(u_{max})}\right)\,,$$

where $U_{max}=\max_j\{U_j\}$, the limit becomes

$$\begin{aligned}\lim_{q\to+\infty}\frac{\tau(q)}{q}&=\lim_{q\to+\infty}-\frac{1}{\ln(b)}\cdot\frac{\ln\left(\Pr\{U_{max}=u_{max}\}\cdot e^{q\cdot\ln(u_{max})}\right)+1}{q}=\\&=\lim_{q\to+\infty}-\frac{1}{\ln(b)}\cdot\frac{\ln\left(\Pr\{U_{max}=u_{max}\}\right)+q\cdot\ln\left(u_{max}\right)+1}{q}=\\&=-\frac{\ln\left(u_{max}\right)}{\ln(b)}=-\log_b\left(u_{max}\right)<\infty\,.\end{aligned}$$

To ascertain that there really is an oblique asymptote, we have to solve this further limit

$$\begin{aligned}\lim_{q\to+\infty}\tau(q)-\left[-\frac{\ln\left(u_{max}\right)}{\ln(b)}\right]\cdot q&=\lim_{q\to+\infty}\tau(q)+\frac{\ln\left(u_{max}\right)}{\ln(b)}\cdot q=\\&=\lim_{q\to+\infty}-\frac{\ln\left(\Pr\{U_{max}=u_{max}\}\cdot e^{q\cdot\ln(u_{max})}\right)+1}{\ln(b)}+\frac{\ln\left(u_{max}\right)}{\ln(b)}\cdot q=\\&=\lim_{q\to+\infty}-\frac{\ln\left(\Pr\{U_{max}=u_{max}\}\right)+q\cdot\ln\left(u_{max}\right)+1}{\ln(b)}+\frac{\ln\left(u_{max}\right)}{\ln(b)}\cdot q=\\&=-\frac{\ln\left(\Pr\{U_{max}=u_{max}\}\right)}{\ln(b)}-\frac{1}{\ln(b)}=\\&=-\log_b\left(\Pr\{U_{max}=u_{max}\}\right)-\log_b(e)<\infty\,.\end{aligned}$$

Thus, for $q \to +\infty$, the scaling function $\tau(q)$ is asymptotic to the straight line with equation

$$a_1(q) = -\log_b(u_{max}) \cdot q - \log_b(\Pr\{U_{max} = u_{max}\}) - \log_b(e)\,.$$

Let we examine the components of the equation: firstly, $-\log_b(e)$ is always negative; secondly, $-\log_b(\Pr\{U_{max} = u_{max}\})$ is always positive (the least $\Pr\{U_{max} = u_{max}\}$, the higher the summand). So the intercept of the asymptote $a_1(q)$ can be both negative or positive, depending on the odd of the event $U_{max} = u_{max}$. Eventually, the rate of growth $-\log_b(u_{max})$ is positive if we are dealing with microcanical measures ($0 \leq U_{max} \leq 1$), but becomes negative if the measure is canonical ($U_{max} \geq 1$).

The same procedure has to be repeated for $q \to -\infty$.

$$\lim_{q\to-\infty} \frac{\tau(q)}{q} = \lim_{q\to-\infty} \frac{-\log_b\left(\sum_{j=0}^{b-1} \Pr\{U_j = u_j\} \cdot u_j^q\right) - 1}{q} =$$
$$= \lim_{q\to-\infty} -\frac{1}{\ln(b)} \cdot \frac{\ln\left(\sum_{j=0}^{b-1} \Pr\{U_j = u_j\} \cdot e^{q\cdot\ln(u_j)}\right) + 1}{q}\,.$$

Because, as $q \to -\infty$

$$\ln\left(\sum_{j=0}^{b-1} \Pr\{U_j = u_j\} \cdot e^{q\cdot\ln(u_j)}\right) \sim \ln\left(\Pr\{U_{min} = u_{min}\} \cdot e^{q\cdot\ln(u_{min})}\right)\,,$$

where $U_{min} = \min_j\{U_j\}$, the limit becomes

$$\lim_{q\to-\infty} \frac{\tau(q)}{q} = \lim_{q\to-\infty} -\frac{1}{\ln(b)} \cdot \frac{\ln\left(\Pr\{U_{min} = u_{min}\} \cdot e^{q\cdot\ln(u_{min})}\right) + 1}{q} =$$
$$= \lim_{q\to+\infty} -\frac{1}{\ln(b)} \cdot \frac{\ln(\Pr\{U_{min} = u_{min}\}) + q\cdot\ln(u_{min}) + 1}{q} =$$
$$= -\frac{\ln(u_{min})}{\ln(b)} = -\log_b(u_{min}) < \infty\,.$$

The limit for the existence of the intercept is

$$\lim_{q\to-\infty} \tau(q) - \left[-\frac{\ln(u_{min})}{\ln(b)}\right] \cdot q = \lim_{q\to-\infty} \tau(q) + \frac{\ln(u_{min})}{\ln(b)} \cdot q =$$
$$= -\log_b(\Pr\{U_{min} = u_{min}\}) - \log_b(e) < \infty\,.$$

Hence, even for $q \to -\infty$, the scaling function $\tau(q)$ has another slant asymptote, that is

$$a_2(q) = -\log_b(u_{min}) \cdot q - \log_b(\Pr\{U_{min} = u_{min}\}) - \log_b(e)\,.$$

Here, the behaviour of the intercept is the same of the previous case. Nevertheless, the slope of the asymptote is always positive, since necessarily $0 \leq U_{min} \leq 1$, both in the microcanonical case and the canonical one.

Then, we summarize all the properties of the scaling function that have been found:

- The point $(1, 0)$ is a zero for $\tau(q)$;
- The point $(0, -1)$ is the intercept of $\tau(q)$;
- If the measure is microcanonical, the function is non-decreasing;
- The function is concave;
- The function is asymptotical linear. Particularly, when $q \to +\infty$

$$\tau(q) \sim -\log_b(u_{max}) \cdot q - \log_b(\Pr\{U_{max} = u_{max}\}) - \log_b(e),$$

and

$$\tau(q) \sim -\log_b(u_{min}) \cdot q - \log_b(\Pr\{U_{min} = u_{min}\}) - \log_b(e),$$

when $q \to -\infty$.

These properties allow us to infer another property of $\tau(q)$ if a canonical measure is involved. Since the function is always concave and also asymptotic, for $q \to +\infty$, to the straight line which has a negative slope, thus it must have a "cap" form, such as $\bigcap$. Moreover, since the function as a zero for $q = 1$, there must exist another one, which is usually addressed as q_{crit}, that is

$$q_{crit} := \{q > 1 : \tau(q) = 0\}$$

which is present, we remark, *only* for canonical measures. This entity has a great impact on the finiteness of the moments of the measure. As a matter of fact Mandelbrot in [37] showed that, for $q > 1$, the moments of the measure are finite if and only if $\tau(q) > 0$. That eventuality occurs only for $1 < q < q_{crit}$.

As a matter of fact, since the q-th moment of a canonical measure is given by

$$\mathbb{E}\Big\{[\mathcal{M}(\Delta t)]^q\Big\} = \mathbb{E}(\Upsilon^q) \cdot \Delta t^{\tau(q)+1},$$

its finiteness may depends only on either $\tau(q)$ or $\mathbb{E}(\Upsilon^q)$. Since, for finite q, $\tau(q)$ is surely finite, hence the "infiniteness" can be achieved only by a particular behaviour of the random variable Υ which is the random mass on the interval $[0, 1]$ (see Section 3.2).

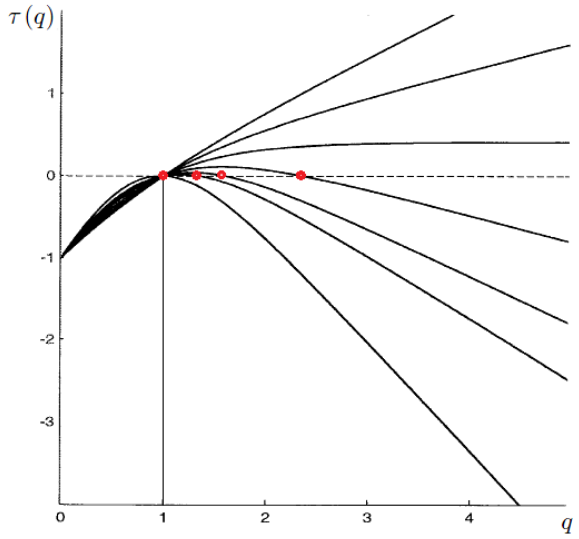


Figure C.1: Possible shapes of $\tau(q)$ for different sequences $\{U_j\}$. Red dots are different q_{crit}.

Indeed, Guivarc'h proved in [21] that Υ has Paretian tails and allows infinite moments, for those values $q \geq q_{crit}$. Thus, the random variable Υ follows a Pareto's distribution of exponent q_{crit}, that is

$$\Pr\{\Upsilon > v\} = \left(\frac{v}{v_{min}}\right)^{-q_{crit}} .$$

C.2 A concise tutorial on Legendre transorm

Definition C.2.1 *Given a concave function $h : \mathbb{R} \to \mathbb{R}$, with $h \in C^2(\mathbb{R})$, we define the* Legendre transform *of $h(x)$ the function*

$$h^\star(\alpha) := \inf_{x \in \mathbb{R}} \{\alpha \cdot x - h(x)\} . \tag{C.1}$$

Hence, we are just measuring the least distance between the line $y = \alpha \cdot x$ and the function $h(x)$. Since the minimum[2] of the parenthetical is reached when

$$\frac{\mathrm{d}\left[\alpha \cdot x - h(x)\right]}{\mathrm{d}x} = 0 ,$$

we find

$$\alpha = h'(x) .$$

Thus α is itself a function of x, and nothing but the fist derivative of $h(x)$. Hence, computing the Legendre transform, we are expressing x as a function

[2]The use of $\inf\{\cdot\}$ instead of $\min\{\cdot\}$ is made in order to be allowed to compute it even for open sets. However, here the difference is negligible.

of α (not vice versa), that is $x(\alpha)$. Moreover, the concavity of h assures the existence of only one minimum/infimum, and thus a consistent definition of the transform itself.

An important property of it is to be an *involution*. In order to prove it, we need two preliminary lemmas.

Lemma C.2.2 *Given $h^\star(\alpha)$ defined as in* (C.1)*, it follows*

$$\left.\frac{\mathrm{d}h^\star(\alpha)}{\mathrm{d}\alpha}\right|_{\alpha=h'(x(\alpha))} = x(\alpha)\,.$$

Proof: Since $h^\star(\alpha) = \alpha\cdot x(\alpha) - h(x(\alpha))$, we find

$$\begin{aligned}\frac{\mathrm{d}h^\star(\alpha)}{\mathrm{d}\alpha} &= \frac{\mathrm{d}\left[\alpha\cdot x(\alpha) - h(x(\alpha))\right]}{\mathrm{d}\alpha} = x(\alpha) + \alpha\cdot x'(\alpha) - h'(x(\alpha))\cdot x'(\alpha) = \\ &= x(\alpha) + \alpha\cdot x'(\alpha) - \alpha\cdot x'(\alpha) = x(\alpha)\,.\end{aligned}$$

□

Lemma C.2.3 *The Legendre transform turns concave functions into concave functions.*

Proof: Thanks to the previous lemma, we have

$$\left.\frac{\mathrm{d}^2h^\star(\alpha)}{\mathrm{d}\alpha^2}\right|_{\alpha=h'(x(\alpha))} = x'(\alpha)\,.$$

Using the previous results, it can be also written as (from now on the subscript notation is avoided)

$$\begin{aligned}\frac{\mathrm{d}^2h^\star(\alpha)}{\mathrm{d}\alpha^2} &= \frac{\mathrm{d}\left[x(\alpha) + \alpha\cdot x'(\alpha) - h'(x(\alpha))\right]}{\mathrm{d}\alpha} = \\ &= x'(\alpha) + \left[x'(\alpha) + \alpha\cdot x''(\alpha)\right] - \left[h''(x(\alpha))\cdot x'(\alpha)^2 + h'(x(\alpha))\cdot x''(\alpha)\right] = \\ &= 2\cdot x'(\alpha) + \alpha\cdot x''(\alpha) - h''(x(\alpha))\cdot x'(\alpha)^2 - \alpha\cdot x''(\alpha) = \\ &= 2\cdot x'(\alpha) - h''(x(\alpha))\cdot x'(\alpha)^2\,.\end{aligned}$$

Comparing with the previous result, we get

$$x'(\alpha) = 2\cdot x'(\alpha) - h''(x(\alpha))\cdot x'(\alpha)^2\,,$$

which is equal to

$$\frac{\mathrm{d}^2h^\star(\alpha)}{\mathrm{d}\alpha^2} = x'(\alpha) = h''(x(\alpha))\cdot x'(\alpha)^2\,.$$

Since h is concave (that is $h''(x) < 0$), we have also

$$\frac{\mathrm{d}^2 h^\star(\alpha)}{\mathrm{d}\alpha^2} < 0\,.$$

□

We are know able to prove the following crucial theorem.

Theorem C.2.4 *The Legendre transform is an* involution*, that is*

$$h(x) = \big(h^\star(\alpha)\big)^\star(x)\,. \tag{C.2}$$

Proof: Let we set $h^\star(\alpha) \equiv f(\alpha)$. Using the definition of Legendre transform, we have

$$f^\star(x) = x \cdot \alpha(x) - f(\alpha(x))\,,$$

where $x = f'(\alpha(x))$. Moreover

$$f(\alpha) = h^\star(\alpha) = \alpha \cdot y(\alpha) - h(y(\alpha))\,,$$

holds true with $\alpha = h'(y(\alpha))$. Using Lemma C.2.2, he have

$$\frac{\mathrm{d}f(\alpha)}{\mathrm{d}\alpha} = f'(\alpha) = y(\alpha)\,.$$

Hence, we have found

$$x = f'(\alpha(x)) = y(\alpha(x))\,.$$

Eventually,

$$\begin{aligned} f^\star(x) = x \cdot \alpha(x) - f(\alpha(x)) = y(\alpha(x)) \cdot \alpha(x) - [\alpha \cdot y(\alpha) - h(y(\alpha))] = \\ = y(\alpha(x)) \cdot \alpha(x) - \alpha(x) \cdot y(\alpha(x)) + h(y(\alpha(x))) = h(y(\alpha(x))) \end{aligned}$$

Making the inverse initial substitution, we find

$$\big(h^\star(\alpha)\big)^\star(x) = h(y(\alpha(x))) = h(x)\,,$$

since $x = y(\alpha(x))$.

□

Furthermore, the following result may be useful.

Theorem C.2.5 (Young's inequality) *Given two concave functions $f, h : \mathbb{R} \to \mathbb{R}$ such that $f(\alpha) = h^\star(\alpha)$, thus*

$$\alpha \cdot x \geq f(\alpha) + h(x)\,.$$

Proof: Using the definition of Legendre transform, and by virtue Theorem C.2.4, we have

$$f(\alpha) = \inf_{x \in \mathbb{R}} \{x \cdot \alpha - h(x)\} \leq x \cdot \alpha - h(x)$$

for all α and x. Thus, the thesis follows.

□

The last results we are interested in discussing are related to many suitable properties of the Legendre transform. Since their proofs are trivial (it is just a matter of computation), we will omit them.

Corollary C.2.6 *Given a concave function $h : \mathbb{R} \to \mathbb{R}$ and a number $k \in \mathbb{R}$, the following properties hold true:*

- *Scaling properties*

$$h(x) = k \cdot g(x) \implies h^\star(\alpha) = k \cdot g^\star\left(\frac{\alpha}{k}\right)$$

and

$$h(x) = g(k \cdot x) \implies h^\star(\alpha) = g^\star\left(\frac{\alpha}{k}\right);$$

- *Translation properties*

$$h(x) = g(x) + k \implies h^\star(\alpha) = g^\star(\alpha) - k$$

and

$$h(x) = g(x + k) \implies h^\star(\alpha) = g^\star(\alpha) - \alpha \cdot k\,;$$

- *Inversion property*

$$h(x) = g^{-1}(x) \implies h^\star(\alpha) = -\alpha \cdot g^\star\left(\frac{1}{\alpha}\right).$$

Now we are able to extend all the previous results on the scaling function $\tau(q)$ and the multifractal spectrum $f(\alpha)$. Since

$$f(\alpha) = \inf_{q \in \mathbb{R}} \{\alpha \cdot q - \tau(q)\},$$

we see that the multifractal spectrum is the Legendre transform of the scaling function. By virtue of Lemma C.2.3, $f''(\alpha) < 0$ that is the multifractal spectrum is a concave function. Moreover, since the Legendre transform is an involution we have

$$f(\alpha) = \tau^\star(\alpha) \qquad \text{and} \qquad \tau(q) = f^\star(q)\,.$$

In addiction, using Young's Inequality, we have

$$\alpha \cdot q \geq f(\alpha) + \tau(q)\,.$$

Here, let we take the explicit expression of the scaling function

$$\tau(q) = -\log_b\left[\mathbb{E}\left(U^q\right)\right] - 1 = -\log_b\left(\sum_{j=0}^{b-1} \Pr\left\{U_j = u_j\right\} \cdot u_j^q\right) - 1\,.$$

Through a point on the graph of coordinates $(q; \tau(q))$, draw the tangent to the graph. We know the tangent's slop is $\alpha(q)$ and its intercept by the ordinate axis is $-f(q)$, that is

$$\alpha(q) = \tau'(q) \qquad \text{and} \qquad -f(q) = \tau(q) - q \cdot \tau'(q)\,.$$

Through the quantities $\alpha(q)$ and $f(q)$, a function $f(\alpha)$ is defined by using q as a parameter. Using Lemma C.2.2, the slope $f'(\alpha)$ is the function $q(\alpha)$ (that is, in other words, the inverse of the function $\alpha(q)$).

Moreover, in Appendix C.1 the following asymptotic relations have been proved

$$\tau(q) \sim -\log_b\left(u_{max}\right) \cdot q - \log_b\left(\Pr\left\{U_{max} = u_{max}\right\}\right) - \log_b(e) \quad \text{as} \quad q \to +\infty$$

and

$$\tau(q) \sim -\log_b\left(u_{min}\right) \cdot q - \log_b\left(\Pr\left\{U_{min} = u_{min}\right\}\right) - \log_b(e) \quad \text{as} \quad q \to -\infty\,.$$

Thus, the slope of the oblique asymptotes are respectively[3]

$$\alpha_{max} = -\log_b\left(u_{min}\right) > 0 \quad \text{and} \quad \alpha_{min} = -\log_b\left(u_{max}\right) \lesseqgtr 0\,.$$

The relation max/min is inverted since the the slope of the asymptote fot $q \to -\infty$ is grater than the one of the asymptote for $q \to +\infty$. Moreover, these two values give the bounds of the support of the multifractal spectrum, that is

$$f : \mathbb{R} \supset [\alpha_{min}, \alpha_{max}] \to \mathbb{R}\,,$$

since they are the least and the highest value α can take. Hence, if the scaling function $\tau(q)$ is defined on the entire real line, its asymptotic linear behaviour implies the spectrum $f(\alpha)$ to be defined only on a closed set of values. Furthermore, using Young's Inequality and the fact that $\tau(0) = -1$ and $\tau(1) = 0$, we can find

$$\alpha \cdot 0 \geq f(\alpha) + \tau(0) \quad \to \quad f(\alpha) \leq 1$$

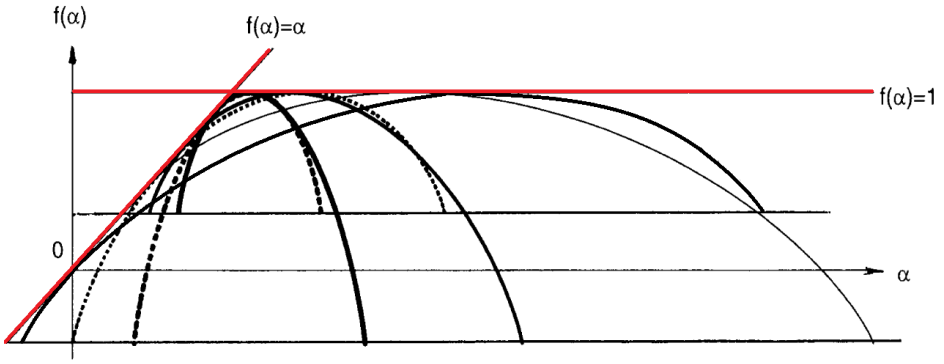


Figure C.2: Possible shapes of $f(\alpha)$ for different $\tau(q)$.

and

$$\alpha \cdot 1 \geq f(\alpha) + \tau(1) \quad \rightarrow \quad f(\alpha) \leq \alpha \,.$$

Figure C.2 illustrates it. About the coordinates of the maximum of $f(\alpha)$, see Appendix C.3.

C.3 Three faces of the Multifractal spectrum

Given the microcanonical measure,

$$\mathcal{M}\left(\Delta t\right) = U_{\eta_1} \cdot U_{\eta_1\eta_2} \cdot \ldots \cdot U_{\eta_1\eta_2\ldots\eta_k}$$

where $\Delta t = b^{-k}$ and t is a b-adic number, let we consider its coarse Hölder exponent

$$\begin{aligned} \alpha_k\left(t\right) &:= \frac{\log \mathcal{M}\left(\left[t, t+b^{-k}\right]\right)}{\log\left(b^{-k}\right)} = \frac{\log_b\left(U_{\eta_1} \cdot U_{\eta_1\eta_2} \cdot \ldots \cdot U_{\eta_1\eta_2\ldots\eta_k}\right)}{\log_b\left(b^{-k}\right)} = \\ &= -\frac{1}{k} \cdot \left[\log_b\left(U_{\eta_1}\right) + \log_b\left(U_{\eta_1\eta_2}\right) + \ldots + \log_b\left(U_{\eta_1\eta_2\ldots\eta_k}\right)\right]. \end{aligned} \tag{C.3}$$

Since the measure is randomly generated because of the multipliers U_{η_1}, $U_{\eta_1\eta_2}, \ldots, U_{\eta_1\eta_2\ldots\eta_k}$, if we choose a fixed interval $[t, t+\Delta t]$, the coarse exponents $\alpha_k(t)$ are identically distributed across the b-adic cells, and hence can be viewed as the realizations of a random variable α_k. In the following, we will use this random variable to study the multifractal spectrum $f(\alpha)$ of the measure $\mathcal{M}$, since it can be directly derived from the asymptotic distribution of α_k.

[3] Since $u_{max} > u_{min}$, we find $\alpha_{max} > \alpha_{min}$. About the sign of α_{min}, see Appendix C.1.

By equation (C.3) and thanks to the way we defined the microcanonical measure, the coarse Hölder exponent is the sample sum of k i.i.d. random variables. Because of it, the distribution of α_k can be analyzed via limit theorems: **(a)** Strong Law of Large Numbers, **(b)** Central Limit Theorem, and **(c)** Large Deviation Theory, which will be then introduced.

(a) Strong Law of large numbers

By the Strong Law of Large Numbers, the sequence of random variable $\{\alpha_k\}_{k\in\mathbb{N}}$ converge almost surely to $\mu_\alpha = \mathbb{E}\big[-\log_b(U)\big]$, that is

$$\Pr\left\{\lim_{k\to\infty}\alpha_k = \mu_\alpha\right\} = 1\,.$$

Moreover, since $\mathbb{E}(U) = \frac{1}{b}$, by Jensen's inequality[4], we have $\mu_\alpha \geq 1$. As k increases, we expect that almost all coarse Hölder exponents are contained in a neighborhood of μ_α. However, the *other* coarse Hölder exponents are also important, and may more than the ones around the mean. As a matter of fact, if we consider (C.3), we can see that the measure can be expressed as

$$\mathcal{M}\left(\left[t, t+b^{-k}\right]\right) = \left(b^{-k}\right)^{\alpha_k(t)} = \Delta t^{\alpha_k(t)}\,,$$

Let Θ denote the set of b-adic intervals with Hölder exponents grater than $\frac{\mu_\alpha+1}{2}$, for large values of k, almost all intervals belong to Θ, but their mass

$$\sum_{t\in\Theta}\mathcal{M}\left(\left[t, t+b^{-k}\right]\right) = \sum_{t\in\Theta}\Delta t^{\alpha_k(t)} \leq b^k\cdot\Delta t^{\frac{\mu_\alpha+1}{2}} = b^{-k\cdot\frac{\mu_\alpha-1}{2}}\,,$$

tends to zero as k goes to infinity. Since the mass must conserve, it has to be in those few b-adic intervals which do not belong to Θ. Since with the Law of Large Numbers we are dealing with centre of the α_k's distribution, information on these "rare events" is presumably contained in the tail of the random variable α_k.

In other words, if α_k differs from μ_α, then $(\alpha_j, \alpha_j+\varepsilon]$, with $\alpha_k \neq \mu_\alpha$ will not contain μ_α for small ε and the chance to observe other coarse exponents which lie in $(\alpha_j, \alpha_j+\varepsilon]$ will decrease exponentially fast with rate given by $f(\alpha)$.

[4]Since $\log_b(\cdot)$ is concave, applying Jensen's inequality to $-\log_b(\cdot)$, we have

$$\mathbb{E}\big[-\log_b(U)\big] \geq -\log_b\big[\mathbb{E}(U)\big] = -\log_b\left(\frac{1}{b}\right) = 1\,.$$

(b) Central Limit Theorem

Assuming that $-\log_b(U)$ has finite variance σ_α^2, we can apply the (Lindeberg-Lévy) Central Limit Theorem, which implies that

$$\sqrt{k} \cdot \frac{\alpha_k - \mu_\alpha}{\sigma_\alpha} \xrightarrow{d} \mathcal{N}(0,1)\,.$$

If $N_k(\alpha)$ stands for the number of coarse Hölder exponents[5] the equal to α_k, in terms of histograms, we have

$$\frac{N_k(\alpha)}{b^k} \sim \frac{1}{\sqrt{2 \cdot \pi \cdot \frac{\sigma_\alpha^2}{k}}} \cdot e^{-\frac{(\alpha-\mu_\alpha)^2}{2 \cdot \sigma_\alpha^2/k}}$$

as $k \to \infty$. After some arrangements and taking the logarithms[6], we found that the multifractal spectrum

$$f(\alpha) \sim 1 - \frac{1}{2 \cdot \ln(b)} \cdot \left(\frac{\alpha - \mu_\alpha}{\sigma_\alpha}\right)^2 \tag{C.4}$$

is locally quadratic around the most probable exponent μ_α.

(c) Large Deviation Theory

Since Large Deviation Theory is not frequently present in textbook on Probability Theory, we give a brief introduction of it. This branch of Probability Theory deals with the decay of the probability of increasingly unlikely

[5] For simplicity, we write $N_k(\alpha)$ instead of $N_k(\alpha, \varepsilon)$. In the following we will set $\varepsilon = \Delta\alpha$.

[6] In fact, as $k \to \infty$

$$N_k(\alpha) \sim \frac{b^k}{\sqrt{2 \cdot \pi \cdot \frac{\sigma_\alpha^2}{k}}} \cdot e^{-\frac{(\alpha-\mu_\alpha)^2}{2 \cdot \sigma_\alpha^2/k}}\,.$$

Taking the logarithms

$$\ln\big[N_k(\alpha)\big] \sim \ln\left(b^k\right) - \ln\left(\sqrt{2 \cdot \pi \cdot \frac{\sigma_\alpha^2}{k}}\right) - \frac{(\alpha - \mu_\alpha)^2}{2 \cdot \frac{\sigma_\alpha^2}{k}}\,,$$

and diving by $\ln\left(b^k\right)$

$$f(\alpha) = \frac{\ln\big[N_k(\alpha)\big]}{\ln\left(b^k\right)} \sim 1 - \frac{\ln\left(\sqrt{2 \cdot \pi \cdot \frac{\sigma_\alpha^2}{k}}\right)}{\ln\left(b^k\right)} - \frac{\frac{k \cdot (\alpha-\mu_\alpha)^2}{2 \cdot \sigma_\alpha^2}}{\ln\left(b^k\right)}\,.$$

Since the second summand goes to zero as $k \to \infty$, we find

$$f(\alpha) \sim 1 - \frac{\frac{k \cdot (\alpha-\mu_\alpha)^2}{2 \cdot \sigma_\alpha^2}}{k \cdot \ln(b)} = 1 - \frac{1}{2 \cdot \ln(b)} \cdot \frac{(\alpha - \mu_\alpha)^2}{\sigma_\alpha^2}\,.$$

events. It was introduced by the Swedish mathematician Harald Cramér, generally more famous for his model on insurance ruin.

Suppose that $X_1, X_2, \ldots$ are independent and identically distributed random variables with mean $\mathbb{E}(X) = \mu_X$ and variance $\mathbb{V}(X) = \sigma_X^2 < +\infty$. If denote the k-th partial sum by

$$S_k := \sum_{i=1}^{k} X_i \,.$$

In order to prove the Strong Law of Large Number, the hypothesis of the Borel-Cantelli Lemma is required to hold, that is

$$\sum_{k=1}^{\infty} \Pr\left\{ \frac{S_k}{k} - \mu_X > \varepsilon \right\} < \infty \,.$$

with $\varepsilon > 0$. However, if such odds are of the order of $\frac{1}{n}$, that is

$$\Pr\left\{ \frac{S_k}{k} - \mu_X > \varepsilon \right\} \approx \frac{1}{n}$$

the Borel-Cantelli Lemma is not applicable anymore, since they are not summable, and the Strong Law of Large Numbers fails. Large Deviation Theory finds out exactly how fast the large deviation probabilities decay. This depends on finer features of the random variable X than merely the finiteness of its variance. Our initial focus is on random variables satisfying

$$\ln\Big[\mathbb{E}\left(e^{q \cdot X}\right)\Big] < \infty$$

where $q \in \mathbb{R}$ and the entire quantity is usually addressed as *cumulant generating function*[7]. In this case the large deviation probabilities decay exponentially and *Cramér's theorem* tells us exactly how fast.

Theorem C.3.1 (**Cramér's theorem**) *Let $X_1, X_2, \ldots$ be independent and identically distributed random variables with mean μ_X, and finite cumulant generating function for all $q \in \mathbb{R}$. Then, we have*

$$\lim_{k\to\infty} \frac{1}{k} \cdot \ln\left[\Pr\left\{ \frac{S_k}{k} > x \right\} \right] = \inf_{q \in \mathbb{R}} \left\{ q \cdot x - \ln\left[\mathbb{E}\left(e^{q \cdot X}\right) \right] \right\}$$

for any $x > \mu_X$, and

$$\lim_{k\to\infty} \frac{1}{k} \cdot \ln\left[\Pr\left\{ \frac{S_k}{k} < x \right\} \right] = \inf_{q \in \mathbb{R}} \left\{ q \cdot x - \ln\left[\mathbb{E}\left(e^{q \cdot X}\right) \right] \right\}$$

for any $x < \mu_X$.

[7] It is just the logarithm of the *moment generating function*.

We can now apply Cramér's theorem to the random variable

$$\alpha_k := -\frac{1}{k} \cdot \left[\log_b (U_{\eta_1}) + \log_b (U_{\eta_1 \eta_2}) + ... + \log_b (U_{\eta_1 \eta_2 ... \eta_k}) \right],$$

that is the k-th partial sum of the i.i.d. random variables $-\log_b(U)$. Thus, we find

$$\lim_{k \to \infty} \frac{1}{k} \cdot \ln \left[\Pr \left\{ \alpha_k > \alpha \right\} \right] = \inf_{q \in \mathbb{R}} \left\{ q \cdot \alpha - \ln \left[\mathbb{E} \left(e^{-q \cdot \log_b (U)} \right) \right] \right\}$$

for any $\alpha > \mu_\alpha$. Calling this quantity as $\delta(\alpha)$, we find[8]

$$\delta(\alpha) = \inf_{q \in \mathbb{R}} \left\{ q \cdot \alpha + \log_b \left[\mathbb{E} \left(U^q \right) \right] \right\} .$$

Since we have set $\tau(q) \equiv -\log_b \left[\mathbb{E} \left(U^q \right) \right] - 1$, the previous expression can be written as

$$\delta(\alpha) = \inf_{q \in \mathbb{R}} \left\{ q \cdot \alpha - \tau(q) \right\} - 1 , \tag{C.5}$$

which shows that $\delta(\alpha) + 1$ is the Legendre transform of $\tau(q)$. Hence the multifractal spectrum is equal to

$$f(\alpha) = \delta(\alpha) + 1 . \tag{C.6}$$

Now, let we come back to the construction of the multifractal spectrum $f(\alpha)$. As we have done before, we subdivide the interval $[0, 1]$ into b^k cells of length $\Delta t = b^{-k}$. Similarly, we can partition the range of α's into intervals of length $\Delta\alpha$, and denote $N_k (\alpha_j)$ the number of coarse Hölder exponents in the interval $(\alpha_j, \alpha_j + \Delta\alpha]$.

As $k \to \infty$, the following heuristic expression

$$\frac{1}{k} \cdot \ln \left[\frac{N_k (\alpha_j)}{b^k} \right] \sim \frac{1}{k} \cdot \ln \big(\Pr \left\{ \alpha_j < \alpha_k \leq \alpha_j + \Delta\alpha \right\} \big) \tag{C.7}$$

[8] Due to independence, it follows

$$\begin{aligned} \delta(\alpha) &= \inf_{q \in \mathbb{R}} \left\{ q \cdot \alpha - \ln \left[\mathbb{E} \left(e^{-q \cdot \log_b (U)} \right) \right] \right\} = \inf_{q \in \mathbb{R}} \left\{ q \cdot \alpha - \ln(b) \cdot \log_b \left[\mathbb{E} \left(e^{-q \cdot \log_b (U)} \right) \right] \right\} = \\ &= \inf_{q \in \mathbb{R}} \left\{ q \cdot \alpha - \log_b \left[\mathbb{E} \left(e^{-q \cdot \log_b (U) \cdot \ln(b)} \right) \right] \right\} = \inf_{q \in \mathbb{R}} \left\{ q \cdot \alpha - \log_b \left[\mathbb{E} \left(b^{-q \cdot \log_b (U)} \right) \right] \right\} = \\ &= \inf_{q \in \mathbb{R}} \left\{ q \cdot \alpha - \log_b \left[\mathbb{E} \left(b^{-\log_b (U^q)} \right) \right] \right\} = \inf_{q \in \mathbb{R}} \left\{ q \cdot \alpha + \log_b \left[\mathbb{E} \left(U^q \right) \right] \right\} . \end{aligned}$$

has to be postulated[9]. Since

$$\begin{aligned}\Pr\{\alpha_j < \alpha_k \leq \alpha_j + \Delta\alpha\} &= \Pr\{\alpha_k \leq \alpha_j + \Delta\alpha\} - \Pr\{\alpha_k \leq \alpha_j\} = \\ &= \Pr\{\alpha_k > \alpha_j\} - \Pr\{\alpha_k > \alpha_j + \Delta\alpha\} = \\ &= \Pr\{\alpha_k > \alpha_j\} \cdot \left(1 - \frac{\Pr\{\alpha_k > \alpha_j + \Delta\alpha\}}{\Pr\{\alpha_k > \alpha_j\}}\right),\end{aligned}$$

expressing in term of tail behaviour, we can apply Large Deviation Theory. Since $\delta(\alpha_j + \Delta\alpha) < \delta(\alpha_j)$, applying (C.5) we find[10]

$$\frac{\Pr\{\alpha_k > \alpha_j + \Delta\alpha\}}{\Pr\{\alpha_k > \alpha_j\}} \sim \left(b^k\right)^{\frac{\delta(\alpha_j+\Delta\alpha)-\delta(\alpha_j)}{\ln(b)}}$$

which hence vanishes as $k \to \infty$. Thus the previous expression becomes

$$\Pr\{\alpha_j < \alpha_k \leq \alpha_j + \Delta\alpha\} \sim \Pr\{\alpha_k > \alpha_j\}$$

as k goes to infinity. Hence (C.7) can be rewritten as

$$\frac{1}{k} \cdot \ln\left[\frac{N_k(\alpha_j)}{b^k}\right] \sim \frac{1}{k} \cdot \ln\big(\Pr\{\alpha_k > \alpha_j\}\big) \sim \delta(\alpha_j)\,, \tag{C.8}$$

for any $\alpha_j > \mu_\alpha$.

A similar procedure could be done starting from the definition of multifractal spectrum (3.13), that is

$$f(\alpha_j) := \lim_{k\to\infty} \frac{\log[N_k(\alpha_j)]}{\log(b^k)}\,.$$

If we choose to use the logarithm with base b, it turns into

$$\begin{aligned}f(\alpha_j) \sim \frac{\log_b[N_k(\alpha_j)]}{\log_b(b^k)} = \frac{1}{k} \cdot \log_b[N_k(\alpha_j)] = \frac{1}{k} \cdot \log_b\left[\frac{N_k(\alpha_j)}{b^k} \cdot b^k\right] = \\ = \frac{1}{k} \cdot \log_b\left[\frac{N_k(\alpha_j)}{b^k}\right] + \frac{1}{k} \cdot \log_b\big(b^k\big) = \frac{1}{k} \cdot \log_b\left[\frac{N_k(\alpha_j)}{b^k}\right] + 1\end{aligned}$$

as $k \to \infty$. Since (C.8) implies also

$$\frac{1}{k} \cdot \log_b\left[\frac{N_k(\alpha_j)}{b^k}\right] \sim \frac{1}{k} \cdot \log_b\big(\Pr\{\alpha_k > \alpha_j\}\big) \sim \delta(\alpha_j)$$

we find the asymptotic equivalence of (C.6)

$$f(\alpha_j) \sim \delta(\alpha_j) + 1\,, \tag{C.9}$$

[9] This expression holds exactly, for any k, for binomial and multinomial measures since the coarse Hölder exponents are discrete. In more general cases, like the microcanonical measure, it has to be postulated.

[10] By Large Deviation Theory we have $\ln[\Pr\{\alpha_k > \alpha\}] \sim k \cdot \delta(\alpha)$, as $k \to \infty$.

confirming Large Deviation Theory. Identical argumentations are valid for $\alpha_j < \mu_\alpha$.

Thus, the Large Deviation Theory approach to multifractal spectrum gives us further information. As a matter of fact, we have shown that $f(\alpha)$ is the limit of

$$\frac{1}{k} \cdot \log_b \left(\Pr \left\{ \alpha_k > \alpha \right\} \right) + 1 \qquad \text{if } \alpha > \mu_\alpha$$

and

$$\frac{1}{k} \cdot \log_b \left(\Pr \left\{ \alpha_k < \alpha \right\} \right) + 1 \qquad \text{if } \alpha < \mu_\alpha \, .$$

We note particularly that $f(\alpha) \leq 1$, and it increases for $\alpha < \mu_\alpha$ and decrease for $\alpha > \mu_\alpha$. Hence $\alpha = \mu_\alpha$ is a maximum and $f(\alpha)$ is a concave function. Morover, since

$$f(\mu_\alpha) = \inf_{q \in \mathbb{R}} \left\{ q \cdot \mu_\alpha - \tau(q) \right\}$$

and it is minimal when $q = 0$, thus $f(\mu_\alpha) = -\tau(0) = 1$, having set the function $\tau(q) \equiv -\log_b \left[\mathbb{E} \left(U^q \right) \right] - 1$. This result receives a simple interpretation in terms of fractal dimension: The set of instants with exponent μ_α has *Lebesgue measure* equal to one (that is what the Strong Law of Large Number states).

We conclude this section remarking that, throughout the dissertation, we have given different notions of multifractal spectrum $f(\alpha)$, and that we have linked it with other quantities by the occasions. Summarizing $f(\alpha)$ can be viewed as:

- The limit of a renormalized histogram of coarse Hölder exponents

$$f(\alpha) = \lim_{\Delta\alpha \to 0} \liminf_{k \to 0} \frac{\log \left[N_k(\alpha, \Delta\alpha) \right]}{\log \left(b^k \right)} \, ;$$

- The fractal dimension of the set of instants with Hölder exponent equal to α

$$f(\alpha) = \dim_{\mathcal{H}} \left(T_\alpha \right) \, ;$$

- The limit provided by Large Deviation Theory

$$f(\alpha) = \begin{cases} \lim_{k \to \infty} \frac{1}{k} \cdot \log_b \left(\Pr \left\{ \alpha_k > \alpha \right\} \right) + 1 & \text{if } \alpha > \mu_\alpha \\ \\ \lim_{k \to \infty} \frac{1}{k} \cdot \log_b \left(\Pr \left\{ \alpha_k < \alpha \right\} \right) + 1 & \text{if } \alpha < \mu_\alpha \end{cases} .$$

However, it is crucial to highlight that these definition *may* coincide. But there may be also discrepancies. For a deep analysis of such an eventuality see [23].

However, the previous explanations are based on microcanonical measure. It will be easy to extend the to the more general canonical case. As a matter of fact, considering the canonical measure, the coarse Hölder exponent is given by

$$\alpha_k(t) := \frac{\log \mathcal{M}\left(\left[t, t+b^{-k}\right]\right)}{\log\left(b^{-k}\right)} = \frac{\log_b\left[\Upsilon \cdot \left(U_{\eta_1} \cdot U_{\eta_1\eta_2} \cdot ... \cdot U_{\eta_1\eta_2...\eta_k}\right)\right]}{\log_b\left(b^{-k}\right)} =$$
$$= -\frac{1}{k} \cdot \log_b(\Upsilon) - \frac{1}{k} \cdot \left[\log_b\left(U_{\eta_1}\right) + \log_b\left(U_{\eta_1\eta_2}\right) + ... + \log_b\left(U_{\eta_1\eta_2...\eta_k}\right)\right].$$

Since when $k \to \infty$ the summand $-\frac{1}{k} \cdot \log_b(\Upsilon) \to 0$, hence the latter does not affect $f(\alpha)$. Then all the previous results on multifractal spectrum are still valid.

Finally, it is straightforward to prove that, if the random quantities $V_k = -\log_b\left(U_{\eta_1\eta_2...\eta_k}\right)$ have Gaussian density, that is $V_k \overset{d}{\sim} \mathcal{N}\left(\mu_\alpha, \sigma_\alpha^2\right)$ and[11]

$$p(v) = \frac{1}{\sqrt{2 \cdot \pi \cdot \sigma_\alpha^2}} \cdot e^{-\frac{(v-\mu_\alpha)^2}{2 \cdot \sigma_\alpha^2}},$$

hence the multifractal spectrum is exactly (C.4).

It implies the random multipliers $U_{\eta_1\eta_2...\eta_k}$ to be log-normal, which implies that $U > 1$ with positive probability. Therefore the conservation of mass cannot at any stage, and then the multiplicative measure is necessarily *canonical*.

Since

$$\alpha_k := -\frac{1}{k} \cdot \left[\log_b\left(U_{\eta_1}\right) + \log_b\left(U_{\eta_1\eta_2}\right) + ... + \log_b\left(U_{\eta_1\eta_2...\eta_k}\right)\right] =$$
$$= -\frac{1}{k} \cdot \sum_{i=1}^{k} V_k,$$

the coarse Hölder exponent has density equal to

$$k \cdot p_k(k \cdot \alpha),$$

where $p_k(\cdot)$ is the density of the k-th convolution product of p. Since the k-th convolution product of independent Gaussian densities is given by

$$p_k(v) := \underbrace{(p * p * ... * p)}_{k \text{ times}}(v) = \frac{1}{\sqrt{2 \cdot \pi \cdot \left(\sum_{i=1}^{k} \sigma_i^2\right)}} \cdot e^{-\frac{(v-\sum_{i=1}^{k} \mu_i)^2}{2 \cdot \sum_{i=1}^{k} \sigma_i^2}},$$

[11] Here, we indicate the p.d.f. of V_k with $p(v)$ rather than with $f(v)$, in order to avoid any confusion with the notation of the multifractal spectrum.

hence the same convolution of i.i.d Gaussian random variables with mean μ_α and variance σ_α^2 is equal to

$$p_k(v) = \frac{1}{\sqrt{2 \cdot \pi \cdot k \cdot \sigma_\alpha^2}} \cdot e^{-\frac{(v-k\cdot\mu_\alpha)^2}{2\cdot k\cdot\sigma_\alpha^2}} .$$

Using Cramér's theorem we now are able to compute the spectrum. In fact

$$\begin{aligned}
\lim_{k\to\infty} \frac{1}{k} \cdot \log_b\left(\Pr\{\alpha_k > \alpha\}\right) &= \lim_{k\to\infty} \frac{1}{k} \cdot \log_b\left[k \cdot \frac{1}{\sqrt{2 \cdot \pi \cdot k \cdot \sigma_\alpha^2}} \cdot e^{-\frac{(k\cdot\alpha - k\cdot\mu_\alpha)^2}{2\cdot k\cdot\sigma_\alpha^2}}\right] = \\
&= \lim_{k\to\infty} \frac{1}{k} \cdot \log_b\left[k \cdot \frac{1}{\sqrt{2 \cdot \pi \cdot k \cdot \sigma_\alpha^2}} \cdot e^{-\frac{k\cdot(\alpha-\mu_\alpha)^2}{2\cdot\sigma_\alpha^2}}\right] = \\
&= \lim_{k\to\infty} \left[\frac{\log_b(k)}{k} - \frac{\log_b(\sqrt{2 \cdot \pi \cdot k \cdot \sigma_\alpha^2})}{k} - \frac{\log_b(e)\cdot(\alpha-\mu_\alpha)^2}{2\cdot\sigma_\alpha^2}\right] = \\
&= -\frac{1}{2\cdot\ln(b)} \cdot \left(\frac{\alpha-\mu_\alpha}{\sigma_\alpha}\right)^2 .
\end{aligned}$$

Hence, since the spectrum is given by $1 + \lim_{k\to\infty} \frac{1}{k} \cdot \log_b\left(\Pr\left\{\alpha_k > \alpha\right\}\right)$, we have found

$$f(\alpha) = 1 - \frac{1}{2\cdot\ln(b)} \cdot \left(\frac{\alpha-\mu_\alpha}{\sigma_\alpha}\right)^2 . \tag{C.10}$$

Moreover, since $\mathbb{E}(U) = \frac{1}{b}$ and $V = -\log_b(U)$, we have

$$\frac{1}{b} = \mathbb{E}\left(b^{-V}\right) = \mathbb{E}\left(e^{-V\cdot\ln(b)}\right) .$$

Furthermore, using the reproductive property of the normal distribution, it ensues

$$-V \cdot \ln(b) \overset{d}{\sim} \mathcal{N}\left[-\mu_\alpha \cdot \ln(b), \sigma_\alpha^2 \cdot \ln(b)^2\right] .$$

Thus, thanks to the properties of log-normal distributions, we are able to compute the previous expectation

$$\mathbb{E}\left(e^{-V\cdot\ln(b)}\right) = e^{-\mu_\alpha\cdot\ln(b) + \frac{\sigma_\alpha^2\cdot\ln(b)^2}{2}} = b^{-\mu_\alpha} \cdot b^{\frac{\ln(b)\cdot\sigma_\alpha^2}{2}} ,$$

and solve

$$b^{-1} = b^{-\mu_\alpha} \cdot b^{\frac{\ln(b)\cdot\sigma_\alpha^2}{2}}$$

for $\ln(b)$. We find

$$\ln(b) = \frac{2\cdot(\mu_\alpha - 1)}{\sigma_\alpha^2} .$$

Substituting it in (C.10), we find an expression for the multifractal spectrum which does not depend on σ_α^2 nor on the base b. That is

$$f(\alpha) = 1 - \frac{(\alpha - \mu_\alpha)^2}{4 \cdot (\mu_\alpha - 1)} \,. \tag{C.11}$$

Appendix D

Appendix

The following code is used to infer from the data the required parameters to simulate an (arithmetic) Brownian motion and a geometric Brownian motion. All the following codes are based on ENI.MI. Of course, in order to get the same results for the other stocks, it is sufficient to substitute ENI.MI with the corresponding name inside the function `get.hist.quote`.

D.1 R codes contained in 4.1

```
#Code for estimating the parameters of a BM
 and to replicate one path via those estimates

library(tseries)

op <- par(mfcol=c(2,1), cex=0.7)

price <- ts(get.hist.quote("ENI.MI",start="2008-01-01",
          end="2013-07-31",quote="Close",compression="d"))
ret <- diff(log(price))

sigma <-(var(ret))^(0.5)
mu <- mean(ret)+sigma^2/2

wn <- rnorm(length(price),mean=0,sd=1)
rw <- cumsum(wn)
time <- 1:length(price)
S_0 <- as.numeric(price[1])

gbm_sim <- ts(S_0*exp((mu-sigma^2/2)*time+sigma*rw))
plot(gbm_sim)
```

```
wn_sim <- ts(diff(log(gbm_sim)))
plot(wn_sim)



#Code for estimating the parameters of a FBM
 and to replicate one path via those estimates

library(tseries)
library(fArma)

op <- par(mfcol=c(2,1), cex=0.7)

price <- get.hist.quote("ENI.MI",start="2008-01-01",
       end="2013-07-31",quote="Close",compression="d")
ret <- diff(log(price))

sigma <-(var(ret))^(0.5)
mu <- mean(ret)+sigma^2/2

lrd <- rsFit(ret, levels=length(ret)/10, minnpts=1,
           doplot=TRUE,cut.off=10^c(0.1,5))
H<-as.numeric(lrd@hurst[1])

fgn <- fgnSim(length(price), H)
frw <- ts(cumsum(fgn))
time <- 1:length(price)
S_0 <- as.numeric(price[1])

gfbm_sim <- ts(S_0*exp((mu-sigma^2/2)*time+sigma*frw))
plot(gfbm_sim)
fgn_sim <- ts(diff(log(gfbm_sim)))
plot(fgn_sim)



#Code for estimating the parameters of a geometric LSM
 and to replicate one path via those estimates

library(tseries)
library(fBasics)
library(stabledist)

op <- par(mfcol=c(2,1), cex=0.7)
```

```
price <- ts(get.hist.quote("ENI.MI",start="2008-01-01",
          end="2013-07-31",quote="Close",compression="d"))
ret <- diff(log(price))

stable <- stableFit(ts(ret),doplot = FALSE)

lsn <- rstable(n = length(price),
       alpha =as.numeric(stable@fit$estimate["alpha"]),
       beta = as.numeric(stable@fit$estimate["beta"]),
       gamma=1, delta=0)
lsm <- ts(cumsum(lsn))
time <- 1:length(price)
S_0 <- as.numeric(price[1])

gamma <- stable@fit$estimate["gamma"]
mu <- mean(ret)+gamma^2

glsm_sim <- ts(S_0*exp((mu-gamma^2)*time+sqrt(2)*gamma*lsm))
plot(glsm_sim)
lsn_sim <- ts(diff(log(glsm_sim)))
plot(lsn_sim)
```

D.2 R codes contained in 4.2

```
#Code for drawing the Q-Q plots of the standardized log-returns

library(tseries)

op <- par(mfcol=c(2,2), cex=0.7)

price <- as.numeric(get.hist.quote("ENI.MI",start="2008-01-01",
                    end="2013-07-31",quote="Close",compression="d"))
ret <- diff(log(price))
ret_sd <-(ret-mean(ret))/sd(ret)

qqnormPlot(ret_sd,rug=FALSE)

price <- as.numeric(get.hist.quote("G.MI",start="2008-01-01",
                    end="2013-07-31",quote="Close",compression="d"))
ret <- diff(log(price))
```

```
ret_sd <-(ret-mean(ret))/sd(ret)

qqnormPlot(ret_sd,rug=FALSE)

price <- as.numeric(get.hist.quote("FNC.MI",start="2008-01-01",
                    end="2013-07-31",quote="Close",compression="d"))
ret <- diff(log(price))
ret_sd <- (ret-mean(ret))/sd(ret)

qqnormPlot(ret_sd,rug=FALSE)

price <- as.numeric(get.hist.quote("UCG.MI",start="2008-01-01",
                    end="2013-07-31",quote="Close",compression="d"))
ret <- diff(log(price))
ret_sd <- (ret-mean(ret))/sd(ret)

qqnormPlot(ret_sd,rug=FALSE)
```

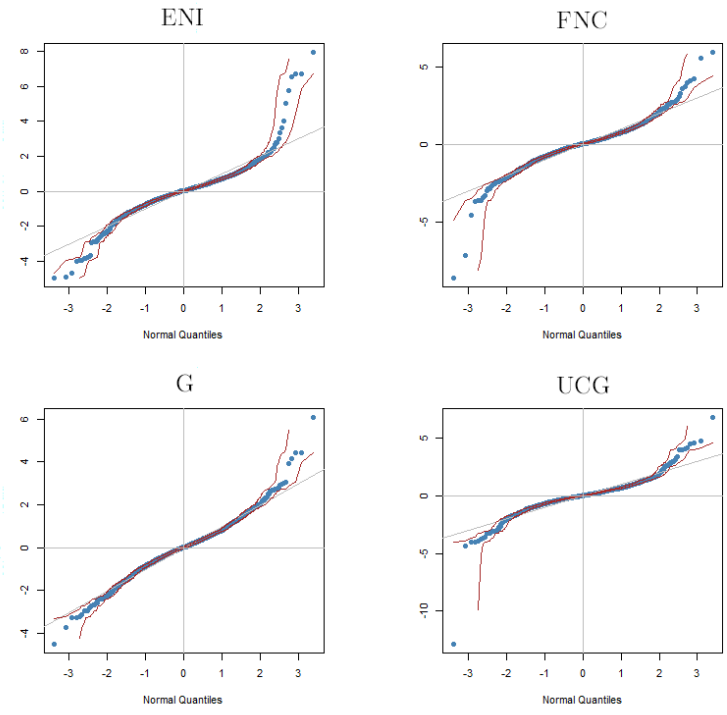


Figure D.1: Q-Q plots of the standardized log-returns.

```
#Code for comparing the estimated densities of standardized
 log-returns with the Gaussian curve

library(tseries)

op <- par(mfcol=c(2,2), cex=0.7)

price <- as.numeric(get.hist.quote("ENI.MI",start="2008-01-01",
                    end="2013-07-31",quote="Close",compression="d"))
ret <- diff(log(price))
ret_sd <-(ret-mean(ret))/sd(ret)

x <- seq(-5,5,by=0.1)
f <- 1/sqrt(2*pi)*exp(-x^2/2)

plot(density(ret_sd), type='h',lwd=3, col="mediumspringgreen")
lines(f~x,type='l',col="black",lwd=2)

price <- as.numeric(get.hist.quote("G.MI",start="2008-01-01",
                  end="2013-07-31",quote="Close",compression="d"))
ret <- diff(log(price))
ret_sd <-(ret-mean(ret))/sd(ret)

plot(density(ret_sd), type='h',lwd=3, col="hotpink")
lines(f~x,type='l',col="black",lwd=2)

price <- as.numeric(get.hist.quote("FNC.MI",start="2008-01-01",
                  end="2013-07-31",quote="Close",compression="d"))
ret <- diff(log(price))
ret_sd <- (ret-mean(ret))/sd(ret)

plot(density(ret_sd), type='h',lwd=3, col="firebrick1")
lines(f~x,type='l',col="black",lwd=2)

price <- as.numeric(get.hist.quote("UCG.MI",start="2008-01-01",
                  end="2013-07-31",quote="Close",compression="d"))
ret <- diff(log(price))
ret_sd <- (ret-mean(ret))/sd(ret)

plot(density(ret_sd), type='h',lwd=3, col="limegreen")
lines(f~x,type='l',col="black",lwd=2)
```

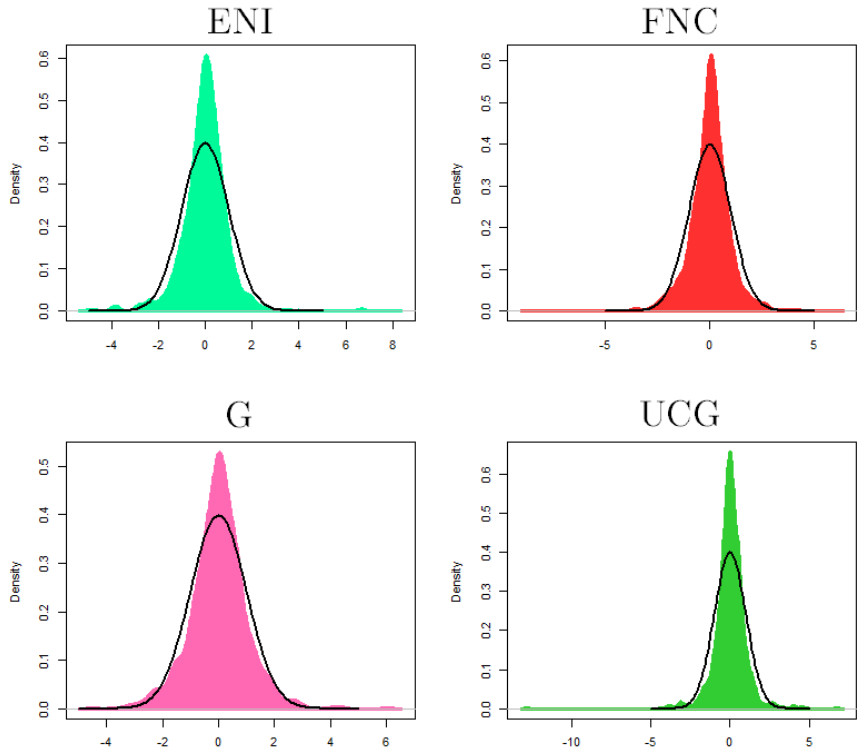


Figure D.2: Densities of the standardized log-returns.

```
#Code for estimating and plotting the scaling function
 and the premultiplicative factor

library(tseries)
price <- ts(get.hist.quote("ENI.MI",start="2008-01-01",
          end="2013-07-31",quote="Close",compression="d"))

op <- par(mfcol=c(1,2), cex=0.7)

t <- c(2,3,4,5,10,15,20)
q <- seq(0.01,10,by=0.01)
y <- c(matrix(0,1,length(t)))
delta <- c(log(t))

tau <- c(matrix(0,1,length(q)))
c <- c(matrix(0,1,length(q)))

r <- 1
for (qq in q){
```

```
    y <- rep(0,length(t))
    s <- 1
    for(time in t){
        x <- seq(1,length(price),by=time)
        y[s] <- log(sum((abs(log(price[x[-1]])
         -log(price[x[-length(x)]])))^qq)) -log(length(price))
        s <- s+1
    }

    lm <- lm(y~delta)

    tau[r] <- as.numeric(lm$coefficients[2])
    c[r] <- exp(as.numeric(lm$coefficients[1]))
    r <- r+1
}

tau <- c(-1,tau)
q <- c(0,q)
c <- c(1,c)
bm <- 0.5*q-1

plot(bm~q,type='l',col="lightblue",lwd=3)
lines(tau~q,type='l',col="purple",lwd=3)
plot(c~q,type='l',col="red",lwd=3)




#Code for estimating the self-affinity index H

library(tseries)
price <- ts(get.hist.quote("ENI.MI",start="2008-01-01",
            end="2013-07-31",quote="Close",compression="d"))

t <- c(2,3,4,5,10,15,20)
q <- seq(0.01,10,by=0.01)
y <- c(matrix(0,1,length(t)))
delta <- c(log(t))

tau <- c(matrix(0,1,length(q)))

r <- 1
for (qq in q){
    y <- rep(0,length(t))
```

```
    s <- 1
    for(time in t){
        x <- seq(1,length(price),by=time)
        y[s] <- log(sum((abs(log(price[x[-1]])
         -log(price[x[-length(x)]])))^qq))-log(length(price))
        s <- s+1
    }

    lm <- lm(y~delta)

    tau[r] <- as.numeric(lm$coefficients[2])
    r <- r+1
}

tau <- c(-1,tau)
q <- c(0,q)

tauf <- approxfun(q,tau)
zero <- uniroot(tauf,c(1.7,2.91))
H <- 1/zero$root



#Code for estimating the minimum admissible alpha

library(tseries)
library(fArma)
price <- ts(get.hist.quote("ENI.MI",start="2008-01-01",
            end="2013-07-31",quote="Close",compression="d"))

t <- c(2,3,4,5,10,15,20)
q <- c(100,99.999)
y <- c(matrix(0,1,length(t)))
delta <- c(log(t))

tau <- c(matrix(0,1,length(q)))

r <- 1
for (qq in q){
    y <- rep(0,length(t))
    s <- 1
    for(time in t){
        x <- seq(1,length(price),by=time)
        y[s] <- log(sum((abs(log(price[x[-1]])
```

```
         -log(price[x[-length(x)]])))^qq)) -log(length(price))
    s <- s+1
    }

    lm <- lm(y~delta)

    tau[r] <- as.numeric(lm$coefficients[2])
    r <- r+1
}

alpha_min <- (tau[1]-tau[2])/(q[1]-q[2])



#Code for estimating the multifractal spectrum via
 Legendre transform, compared with estimated spectrum
 via (C.11)

library(tseries)
price <- ts(get.hist.quote("ENI.MI",start="2008-01-01",
            end="2013-07-31",quote="Close",compression="d"))

t <- c(2,3,4,5,10,15,20)
q <- c(100,99.999)
y <- c(matrix(0,1,length(t)))
delta <- c(log(t))

tau <- c(matrix(0,1,length(q)))

r <- 1
for (qq in q){
    y <- rep(0,length(t))
    s <- 1
    for(time in t){
        x <- seq(1,length(price),by=time)
        y[s] <- log(sum((abs(log(price[x[-1]])
         -log(price[x[-length(x)]])))^qq)) -log(length(price))
        s <- s+1
    }

    lm <- lm(y~delta)

    tau[r] <- as.numeric(lm$coefficients[2])
    r <- r+1
```

```
}

alpha_min <- (tau[1]-tau[2])/(q[1]-q[2])

q <- seq(0.01,10,by=0.01)

tau <- c(matrix(0,1,length(q)))

r <- 1
for (qq in q){
    y <- rep(0,length(t))
    s <- 1
    for(time in t){
        x <- seq(1,length(price),by=time)
        y[s] <- log(sum((abs(log(price[x[-1]])
         -log(price[x[-length(x)]])))^qq))-log(length(price))
        s <- s+1
    }

    lm <- lm(y~delta)

    tau[r] <- as.numeric(lm$coefficients[2])
    r <- r+1
}

tau <- c(-1,tau)
q <- c(0,q)

tauf <- approxfun(q,tau)
zero <- uniroot(tauf,c(1.7,2.91))
H <- 1/zero$root

legendre <- function(alpha,tauf){
            F <- function(alpha,x) alpha*x-tauf(x)
            res <- optimize(F,c(q[1],q[length(q)]),alpha=alpha)
            xs <- res$minimum
            alpha*xs-tauf(xs)
}

alpha <- seq(alpha_min,1.1,by=0.001)
spectr <- sapply(alpha,legendre,function(x)+tauf(x))

m <- alpha[which(spectr<0.99)[length(which(spectr<0.99))]]
```

```
alpha <- seq(alpha_min,m,len=15)
spectr <- sapply(alpha,legendre,function(x)+tauf(x))
plot(alpha,spectr)

f <- function(x) 1-(x-m)^2/(4*H*(m-H))
ff <- sapply(alpha,f)
lines(ff~alpha, type='l', col='darkturquoise', lwd=2)


#Code for estimating the parameters of the MMAR
 and to replicate one path via those estimates

library(tseries)
library(fArma)

op <- par(mfcol=c(2,1), cex=0.7)

price <- ts(get.hist.quote("ENI.MI",start="2008-01-01",
            end="2013-07-31",quote="Close",compression="d"))

t <- c(2,3,4,5,10,15,20)
q <- c(100,99.999)
y <- c(matrix(0,1,length(t)))
delta <- c(log(t))

tau <- c(matrix(0,1,length(q)))

r <- 1
for (qq in q){
    y <- rep(0,length(t))
    s <- 1
    for(time in t){
        x <- seq(1,length(price),by=time)
        y[s] <- log(sum((abs(log(price[x[-1]])
         -log(price[x[-length(x)]])))^qq)) -log(length(price))
        s <- s+1
    }

    lm <- lm(y~delta)

    tau[r] <- as.numeric(lm$coefficients[2])
    r <- r+1
}
```

```
alpha_min <- (tau[1]-tau[2])/(q[1]-q[2])

q <- seq(0.01,10,by=0.01)
tau <-c(matrix(0,1,length(q)))

r <- 1
for (qq in q){
    y <- rep(0,length(t))
    s <- 1
    for(time in t){
        x <- seq(1,length(price),by=time)
        y[s] <- log(sum((abs(log(price[x[-1]])
         -log(price[x[-length(x)]])))^qq))-log(length(price))
        s <- s+1
    }

    lm <- lm(y~delta)

    tau[r] <- as.numeric(lm$coefficients[2])
    r <- r+1
}

tau <- c(-1,tau)
q <- c(0,q)

tauf <- approxfun(q,tau)
zero <- uniroot(tauf,c(1.7,2.91))
H <- 1/zero$root

legendre <- function(alpha,tauf){
            F <- function(alpha,x) alpha*x-tauf(x)
            res <- optimize(F,c(q[1],q[length(q)]),alpha=alpha)
            xs <- res$minimum
            alpha*xs-tauf(xs)
}

alpha <- seq(alpha_min,1.1,by=0.001)
spectr <- sapply(alpha,legendre,function(x)+tauf(x))

m <- alpha[which(spectr<0.99)[length(which(spectr<0.99))]]

b <- 2
k <- 1:11
```

```
vec <- function(A) as.vector(t(A))

mu <- m/H
sigma <- sqrt((2*(mu-1))/log(b))
U <- matrix(0,length(k),b)

for (kk in k){
    U[kk,] <- b^(-rnorm(n=b,mu,sigma))}
    m1 <- matrix(U[1,],b,1)
    for(i in 2:k[length(k)]){
        m2 <- NULL
        for(j in as.vector(t(m1))){
        m2 <- rbind(m2,j*U[i,])
    }
    m1 <- m2
}

theta <- vec(m2)

ret <- diff(log(price))
sigma_ret <- (var(ret))^(0.5)
S_0 <- as.numeric(price[1])

x <- cumsum(sqrt(b^k[length(k)]*theta[1:length(price)])
            *sigma_ret*fgnSim(length(price), H))

gmmar_sim <- ts(S_0*exp(x))
plot(gmmar_sim)
nmarr_sim <- ts(diff(log(gmmar_sim)))
plot(nmarr_sim)
```

D.3 R codes contained in 4.3

```
#Code to calculate the price of a call
 via the Black-Scholes formula

library(tseries)

price<-ts(get.hist.quote("ENI.MI",start="2013-03-01",
        end="2013-09-20",quote="Close",compression="d"))
ret<-diff(log(price))
```

```
sigma<-(var(ret))^(0.5)
r<-0.00003
S_0<-as.numeric(price[length(price)])
K<-18
days<-45

d_1<-(log(S_0/K)+(r+sigma^2/2)*days)/(sigma*sqrt(days))
d_2<-d_1-sigma*sqrt(days)
call_BS<-round(S_0*pnorm(d_1)-K*exp(-r*days)*pnorm(d_2),4)
print(call_BS,round=4)



#Monte Carlo simulation of geometric BMs, and
 pricing of a call using the generated paths

library(tseries)

price <- ts(get.hist.quote("ENI.MI",start="2013-03-01",
          end="2013-09-20",quote="Close",compression="d"))
ret <- diff(log(price))

sigma <- (var(ret))^(0.5)
r <- 0.00003
S_0 <- as.numeric(price[length(price)])
K <- 18

days <- 45
paths <- 500000
time <- 1:days

m <- matrix(0,days,paths)
time <- 1:days
for(sim in 1:paths) {
    temp <- cumsum(ts(rnorm(days,mean=0,sd=1)))
    m[,sim] <- S_0*exp((r-sigma^2/2)*time+sigma*temp)
    }

start <- matrix(S_0,1,paths)
monte <- rbind(start,m)
BM_MC <- monte[,1:300]

matplot(BM_MC,type='l')
```

```
payoffs <- 0.5*((m[days,]-K)+abs(K-m[days,]))
call <- round(exp(-r*days)*mean(payoffs),4)
print(call)


#Monte Carlo simulation of geometric FBMs, and
 pricing of a call the generated paths

library(tseries)
library(fArma)

price <- ts(get.hist.quote("ENI.MI",start="2013-03-01",
            end="2013-09-20",quote="Close",compression="d"))
ret <- diff(log(price))

sigma <- (var(ret))^(0.5)
r <- 0.00003
S_0 <- as.numeric(price[length(price)])
K <- 18

lrd <- rsFit(ret, levels=length(ret)/4, minnpts=1,
            doplot=TRUE, cut.off=10^c(0.1,5))
H <- as.numeric(lrd@hurst[1])

days <- 45
paths <- 500000
time <- 1:days

m <- matrix(0,days,paths)
time <- 1:days
for(sim in 1:paths) {
    temp <- ts(cumsum(fgnSim(days, H)))
    m[,sim] <- S_0*exp((r-sigma^2/2)*time+sigma*temp)
    }

start <- matrix(S_0,1,paths)
monte <- rbind(start,m)
FBM_MC <- monte[,1:300]

matplot(FBM_MC,type='l')

payoffs <- 0.5*((m[days,]-K)+abs(K-m[days,]))
call <- round(exp(-r*days)*mean(payoffs),4)
```

```
print(call)


#Monte Carlo simulation of geometric LSMs, and
 pricing of a call using the generated paths

library(tseries)
library(fBasics)
library(stabledist)

price <- ts(get.hist.quote("ENI.MI",start="2013-03-01",
            end="2013-09-20",quote="Close",compression="d"))
ret <- diff(log(price))

r <- 0.00003
sigma <- var(ret)^0.5
S_0 <- as.numeric(price[length(price)])
K <- 18

stable <- stableFit(ts(ret),doplot = FALSE)
gamma <- as.numeric(stable@fit$estimate["gamma"])

days <- 45
paths <- 500000
time <- 1:days

m <- matrix(0,days,paths)
time <- 1:days
for(sim in 1:paths) {
    temp <- cumsum(ts(rstable(n = days,
            alpha =as.numeric(stable@fit$estimate["alpha"]),
            beta = as.numeric(stable@fit$estimate["beta"]),
            gamma=1, delta=0)))
    m[,sim] <- S_0*exp((r-gamma^2)*time+sqrt(2)*gamma*temp)
    }

start <- matrix(S_0,1,paths)
monte <- rbind(start,m)
LSM_MC <- monte[,1:300]

for(i in 1:paths){
    if(m[days,i]>4*S_0)
    m[days,i] <- S_0
    }  #this part of the code is used to avoid infinite prices
```

```
matplot(LSM_MC,type='l')

payoffs <- 0.5*((m[days,]-K)+abs(K-m[days,]))
call <- round(exp(-r*days)*mean(payoffs),4)
print(call)



#Monte Carlo simulation of geometric MMARs, and
 pricing of a call using the generated paths

library(tseries)
library(fArma)

price <- ts(get.hist.quote("ENI.MI",start="2008-01-01",
            end="2013-09-20",quote="Close",compression="d"))
ret <- diff(log(price))

t <- c(2,3,4,5,10,15,20)
q <- c(100,99.999)
y <- c(matrix(0,1,length(t)))
delta <- c(log(t))


tau <- c(matrix(0,1,length(q)))

r <- 1
for (qq in q){
    y <- rep(0,length(t))
    s <- 1
    for(time in t){
        x <- seq(1,length(price),by=time)
        y[s] <- log(sum((abs(log(price[x[-1]])
         -log(price[x[-length(x)]])))^qq)) -log(length(price))
        s <- s+1
    }

    lm <- lm(y~delta)

    tau[r] <- as.numeric(lm$coefficients[2])
    r <- r+1

}
```

```
alpha_min <- (tau[1]-tau[2])/(q[1]-q[2])

q <- seq(0.01,10,by=0.01)
tau <- c(matrix(0,1,length(q)))

r <- 1
for (qq in q){
    y <- rep(0,length(t))
    s <- 1
    for(time in t){
        x <- seq(1,length(price),by=time)
        y[s]<- log(sum((abs(log(price[x[-1]])
         -log(price[x[-length(x)]])))^qq))-log(length(price))
        s <- s+1
    }

    lm<-lm(y~delta)

    tau[r]<-as.numeric(lm$coefficients[2])
    r <- r+1
}

tau <- c(-1,tau)
q <- c(0,q)

tauf <- approxfun(q,tau)

zero <- uniroot(tauf,c(1.7,3))
H <- 1/zero$root

legendre <- function(alpha,tauf){
            F <- function(alpha,x) alpha*x-tauf(x)
            res <- optimize(F,c(q[1],q[length(q)]),alpha=alpha)
            xs <- res$minimum
            alpha*xs-tauf(xs)
            }

alpha <- seq(alpha_min,1.1,by=0.001)
spectr <- sapply(alpha,legendre,function(x)+tauf(x))

m <- alpha[which(spectr<0.99)[length(which(spectr<0.99))]]

sigma_ret <- (var(ret))^(0.5)
```

```
r <-0.00003
S_0 <- as.numeric(price[length(price)])
K <- 18

days <- 45

b <- 2
k<-1:6
numsim <- 500000

bigm <- matrix(0,b^k[length(k)],numsim)
vec <- function(A) as.vector(t(A))

for(sim in 1:numsim){
    mu <- m/H
    sigma <- sqrt((2*(mu-1))/log(b))
    U <- matrix(0,length(k),b)

        for (kk in k){
        U[kk,] <- b^(-rnorm(n=b,mu,sigma))
        }

    m1 <- matrix(U[1,],b,1)

        for(i in 2:k[length(k)]){
        m2 <- NULL
            for(j in as.vector(t(m1))){
            m2 <- rbind(m2,j*U[i,])
            }
    m1 <- m2
        }

bigm[,sim] <- vec(m2)

}

x <- matrix(0,dim(bigm)[1],dim(bigm)[2])

for(i in 1:dim(bigm)[2]){
    x[,i] <- cumsum(sqrt(b^k[length(k)]*as.vector(bigm[,i]))*sigma_ret*
                    fgnSim(b^k[length(k)], 0.5))
    }

x <- x[1:days,]
```

```
M <- ts(S_0*exp(x))

start <- matrix(S_0,1,numsim)
monte <- rbind(start,M)
MMAR_MC <- monte[,1:300]

matplot(MMAR_MC,type='l')

payoffs <- 0.5*((M[days,]-K)+abs(K-M[days,]))
call <- round(exp(-r*days)*mean(payoffs),4)
print(call)
```

Bibliography

[1] Andreoli A., Caravenna F., Dai Pra P. and Posta G. (2012), *Scaling and Multiscaling in Financial Series: A Simple Model*, Advances in Applied Probability, Vol. 44, No. 4, pp. 1018-1051.

[2] Baldovin F. and Stella A. L. (2007), *Scaling and Efficiency Determine the Irreversible Evolution of a Market*, PNAS, Vol. 104, No. 50, 19741-19744.

[3] Baldovin F. and Stella A. L. (2008), *Role of Scaling in the Statistical Modelling of Finance*, Pramana Journal of Physics, Vol. 71, No. 2, pp. 341-352.

[4] Barnsley M. (2012), *Fractals Everywhere*, Dover Publications.

[5] Bianchi M. L., Fabozzi F. J., Kim Y. S. and Rachev S. T. (2011) *Financial Models with Levy Processes and Volatility Clustering*, Wiley.

[6] Bjiork T. (1998), *Arbitrage Theory in Continuous Time*, Oxford University Press.

[7] Calvet L., Fisher A. and Mandelbrot B. (1996), *A Multifractal Model of Asset Returns*, Cowles Foundation for Research in Economics, Discussion Paper No. 1164.

[8] Calvet L., Fisher A. and Mandelbrot B. (1996), *Large Deviations and the Distribution of Price Changes*, Cowles Foundation for Research in Economics, Discussion Paper No. 1165.

[9] Calvet L., Fisher A. and Mandelbrot B. (1996), *Multifractality of Deutschemark/US Dollar Exchange Rates*, Cowles Foundation for Research in Economics, Discussion Paper No. 1165.

[10] Calvet L. and Fisher A. (2012), *Extreme Risk and Fractal Regularity in Finance*, Working Papers Series.

[11] Crow E. L. and Holt D.R., (1973), *Tables and graphs of the stable probability density functions*, J. Res. National Bureau Standards, Vol. 77B, pp. 143-198.

[12] Doob J. L. (1990), *Stochastic Processes*, Wiley-Interscience.

[13] Edgar G. (2007), *Measure, Topology, and Fractal Geometry*, Springer.

[14] Falconer K. (2003), *Fractal Geometry: Mathematical Foundations and Applications*, Wiley.

[15] Fama E. F. (1965), *The Behavior of Stock-Market Prices*, The Journal of Business, Vol. 38, No. 1.

[16] Fama E. F. and Roll R. (1971), *Parameter estimates for symmetric stable distributions*, Journal of the American Statistical Association, Vol. 66, pp. 331-338.

[17] Fama E. F. and Miller H. M. (1972), *The Theory of Finance*, Holt Rinehart & Winston.

[18] Feller W. (1968), *An Introduction to Probability Theory and Its Applications*, Vol.1, Wiley.

[19] Feller W. (1971), *An Introduction to Probability Theory and Its Applications*, Vol.2, Wiley.

[20] Gnedenko B. V. and Kolmogorov A. N. (1968) *Limit Distributions for Sums of Independent Random Variables*, Addison-Wesley.

[21] Guivarc'h Y. (1987), *Remarques sur les solutions d'une équation fonctionnelle non linéaire de B. Mandelbrot*, C.R.A.S., Série I , pp. 139-141.

[22] Kolmogorov A. N. (1940), *Wienersche spiralen und einige andere interessante kurven im Hilbertschen raum*, Comptes Rendus de l'Academie des Sciences de l'URSS, Vol. 26, pp. 115–118.

[23] Mandelbrot B. (1989), *Multifractal Measures, Expecially fot the Geophysicist*, Pure and Applied Geophysics , Vol.131, pp 5-42.

[24] Mandelbrot B. (1999), *Survey of Multifractality in Finance*, Cowles Fundation for Research in Economics, Discussion Paper No. 1238.

[25] Mandelbrot B. (1999), *Multifractals and 1/f Noise: Wild Self-Affinity in Physics*, Springer.

[26] Mandelbrot B. (2001), *Gaussian Self-Affinity and Fractals*, Springer.

[27] Mandelbrot B. (2010), *Fractals and Scaling in Finance: Discontinuity, Concentration, Risk*, Springer.

[28] Mandelbrot B. (2006), *The Misbehavior of Markets: A Fractal View of Financial Turbulence*, Basic Books.

[29] Marcus M. B. (1968), *Hölder conditions for Gaussian processes with stationary increments*, Transactions of the American Mathematical Society, Vol. 134, No. 1, pp. 29-52.

[30] McCulloch J. H. (1986), *Simple consistent estimators of stable distribution parameters*, Communications in Statistics-Simulations, Vol. 15, pp. 1109-1136.

[31] Merton R. C. (1992), *Continuous-Time Finance*, Wiley-Blackwell.

[32] Mörtens P. and Peres Y. (2010), *Brownian Motion*, Cambridge University Press.

[33] Nolan J. P. (2013), *Stable Distributions - Models for Heavy Tailed Data*, Birkhauser.

[34] Øksendal B. K. (2003), *Stochastic Differential Equations: An Introduction with Applications*, Springer.

[35] Pagani C. D. and Salsa S. (1990), *Analisi Matematica: Volume 1*, Zanichelli.

[36] Pagani C. D. and Salsa S. (1991), *Analisi Matematica: Volume 2*, Zanichelli.

[37] Rachev S. T. (2003), *Handbook of Heavy Tailed Distributions in Finance*, North Holland.

[38] Samorodnitsky G. and Taqqu M. S. (1994), *Stable non-Gaussian random processes*, Chapman and Hall.

[39] Sato K. (1999), *Lévy Processes and Infinitely Divisible Distributions*, Cambridge University Press.

[40] Uchaikin V. V. and Zolotarev V. M. (1999), *Chance and Stability, Stable Distributions and their Applications*, VSP International Science Publishers.

[41] Wiener N. (1923), *Differential space*, Journal of Mathematical Physics, Vol. 2, pp. 131-174

[42] Yaglom A. M. (1955), *Correlation theory of processes with random stationary n-th increments*, Transactions of the American Mathematical Society, 37(79):1, pp. 141–196